UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Combinação de Mineração de Dados e Agentes Inteligentes para a Prevenção e Combate ao Crime de Branqueamento de Capitais

"Documento Definitivo"

Doutoramento em Informática

Especialidade de Engenharia Informática

Claudio Reginaldo Alexandre

Tese orientada por:

João Carlos Balsa da Silva, Ph.D.

Documento especialmente elaborado para a obtenção do grau de doutor

2023

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Combinação de Mineração de Dados e Agentes Inteligentes para a Prevenção e Combate ao Crime de Branqueamento de Capitais

Doutoramento em Informática

Especialidade de Engenharia Informática

Claudio Reginaldo Alexandre

Tese orientada por:

João Carlos Balsa da Silva, Ph.D.

Júri:

Presidente:

- Doutor Manuel João Caneira Monteiro da Fonseca, Prof. Associado e Presidente do Departamento de Informática da Faculdade de Ciências da Universidade de Lisboa

Vogais:

- Doutor Jaime Simão Sichman, Prof. Titular da Escola Politécnica da Universidade de São Paulo, Brasil
- Doutor Luís Paulo Gonçalves dos Reis, Prof. Associado com Agregação da Faculdade de Engenharia da Universidade do Porto
- Doutor Sérgio Miguel CarneiroMoro, Prof. Associado com Agregação da Escola de Tecnologia e Arquitetura do ISCTE - Instituto Universitário de Lisboa
- Doutor Luís Miguel Parreira e Correia, Prof. Catedrático da Faculdade de Ciências da Universidade de Lisboa
- Doutora Ana Paula Pereira Afonso, Profa. Auxiliar da Faculdade de Ciências da Universidade de Lisboa
- Doutor João Carlos Balsa da Silva, Prof. Auxiliar da Faculdade de Ciências da Universidade de Lisboa (orientador)

Documento especialmente elaborado para a obtenção do grau de doutor

2023

Agradecimentos

Os desafios enfrentados e as situações vividas no desenvolvimento de qualquer trabalho impõe muita dedicação ao autor e a necessidade do apoio de inúmeras pessoas. Em um doutoramento esta dedicação e apoio são levados a limites extremos, principalmente se consideradas as condições que permearam este trabalho. Mesmo atuando na vida académica como professor e na vida profissional lidando com o tema em estudo, retornar a Universidade como aluno, após tantos anos, foi um grande desafio. Desde a convivência em sala de aula com colegas que tinham a idades dos meus filhos até a corrida para nivelar conhecimento em torno da utilização de ferramentas a serem utilizadas nas disciplinas e no desenvolvimento do trabalho.

Por isso, nesta trajetória tenho muito que agradecer, assim como tenho certeza que, por lapso de memória, deixarei de citar alguém, porém, esclareço que a ordem apresentada a seguir não classifica a importância da participação de cada citada, trata-se apenas de um artifício cronológico.

Meu primeiro agradecimento é para a Instituição Financeira onde trabalho, que financiou esta pesquisa e permitiu meu deslocamento e dedicação exclusiva para o desenvolvimento do trabalho. Importante ressaltar que mesmo antes da defesa desta tese, parte do trabalho foi implementado e está em uso na instituição.

Ainda relacionado com a minha liberação, preciso destacar e fazer um agradecimento especial para o amigo Nelson de Souza, então Diretor da Instituição, que incentivou e defendeu a realização deste trabalho junto a Diretoria Executiva.

O agradecimento mais efusivo precisa ser dirigido ao Prof. João Balsa, orientador deste trabalho, pela paciência, amizade e companheirismo, mantendo sempre sua autoridade e posição de mentoria, mas continuamente compartilhando conhecimento e sabedoria. Paciência que foi amplamente testada com minhas solicitações para atender a lógica, por vezes infernal, da burocracia que permeia o processo de empregado de empresa pública estudando no exterior. Sua disponibilidade, apoio e busca por um nível maior de qualidade foram inestimáveis.

Tenho certeza que o Prof. Helder Coelho não tem noção da importância dele nesta trajetória, ele foi o primeiro professor da Faculdade de Ciências com quem falei quando ainda estava analisando as possibilidades de fazer o doutoramento. Depois ele também foi um dos primeiros para quem mostramos o esboço do trabalho, seguindo sugestão do Prof. Balsa. Sou muito grato pela amizade, por nossas conversas no almoço, envolvendo a política brasileira e portuguesa. Tudo além de comentários sobre o trabalho.

Recorri com dúvidas e pedido de orientação sobre assuntos que auxiliaram o trabalho à Profa. Ana Paula Afonso e ao Prof. Luis Moniz a quem sou muito grato pela cordialidade com que me atenderam. A Profa. Antónia Lopes agradeço pela condução e sugestões na qualificação e pelas orientações nas questões académicas.

Os pedidos foram muitos, as dúvidas mais ainda e sempre tive um atendimento excelente na Área de Estudos Pós-Graduados e a maioria dos atendimentos foram realizados pela Sra. Carla Romero, na pessoa dela agradeço a todos que trabalham neste setor.

Este penúltimo, mas, certamente, o mais profundo agradecimento reservo para minha família: minha esposa Telma, meus filho Ilo e Ives e minha Nora Caroline. Acompanharam-me no período de estudo presencial, suportaram meus momentos de estreasse, compartilharam angústias, contribuíram para decisões e apresentaram sempre o incentivo necessário. Por tudo que passamos, vivemos, aproveitamos e aprendemos, classifico o período vivido em Lisboa como o mais feliz e profícuo da minha vida.

Por último, porém, o mais importante de todos por cobrir todo o espectro de acontecimentos, reservo a minha crença e, por isso, agradeço a Deus por tudo que permitiu, e propiciou.

Postumamente dedico ao meu pai Francisco Carneiro e minha mãe Lucia Reginaldo,
uma pequena recompensa por todo o sacrifício.

Resumo

As últimas décadas foram marcadas por dois fatos importantes: organizações de combate ao crime de Branqueamento de Capitais (BC) foram criadas; e os Bancos Centrais, normalmente responsáveis pelo controle e definição de normas, ampliaram as leis de Anti-Branqueamento de Capitais (ABC). A agilidade na adaptação do *modus operandi* dos fraudadores e a falta de informação sistematizada que associe os clientes suspeitos comunicados com a comprovação do crime (classificado neste trabalho como um risco sistémico) são fatores que dificultam a automatização do processo de ABC e pode explicar as raras publicações com soluções inovadoras. A abordagem apresentada nesta tese modifica o tratamento genérico utilizado pela maioria dos trabalhos publicados, indo além da sinalização de transações suspeitas, auxiliando o Analista de ABC na tomada de decisão. Esta tese apresenta uma forma inovadora de integração de processos de aprendizagem, numa perspetiva de grupos de risco, com agentes auxiliando a análise e a tomada de decisão, resultando na implementação do sistema multi-agente denominado Jano, que levou a uma melhoria clara nos resultados relativos à identificação e sinalização de clientes suspeitos de BC. Para cada cliente foi gerado um perfil representando seu padrão de comportamento transacional, permitindo a adoção de uma abordagem que identifica e classifica o nível de risco de BC de cada perfil. Os dados utilizados referem-se a dois anos de movimentações, um com 30 e outro com 32 milhões de transações relevantes. De modo a avaliar a relevância das propostas apresentadas, seis meses de transações foram utilizadas para o teste final, um conjunto dos perfis sinalizados foi submetido aos Analisas de ABC da instituição financeira (IF) que financiou a pesquisa e o resultado foi comparado com outros sistemas em uso naquela IF. Os números obtidos nas métricas F1-score e *Matthews Correlation Coefficient* permitem concluir que com a metodologia proposta nesta tese foi possível obter resultados melhores, destacando que 76% dos perfis confirmados como suspeitos não foram reportados ao órgão regulador na época da sua ocorrência, porque nenhum sistema em utilização na IF sinalizou-os como suspeitos.

Palavras-chave: Anti-Branqueamento de Capitais; Sistema Multi-agente; Agentes Inteligentes; Aprendizado de Máquina; Avaliação de Risco

Abstract

The last decades have been marked by two important facts: organizations to combat Money Laundering (ML) have been created, and Central Banks, usually responsible for control and definition of norms, have expanded Anti-Money Laundering (AML) laws. The agility in the adaptation of the fraudsters' *modus operandi* and the lack of systematic information that associates the suspicious clients communicated with the proof of the crime (classified in this work as a systemic risk), are factors that hinder the automation of the AML process and can explain the rare publications with innovative solutions. The approach presented in this thesis modifies the generic treatment used by most published works, going beyond the signaling of suspicious transactions, assisting the AML Analyst in decision making. This thesis presents an innovative way of integrating learning processes, from a risk groups perspective, with agents helping the analysis and decision making, resulting in the implementation of the multi-agent system called Jano, which has led to a clear improvement in the results related to the identification and signaling of ML suspicious clients. For each client, a profile representing their transactional behavior pattern has been generated, allowing the adoption of an approach that identifies and classifies the ML risk level of each profile. The data used refer to two years of transactions, with an average of 32 million relevant transactions. To evaluate the relevance of the proposals presented, six months of transactions were used for the final test, a set of signaled profiles was submitted to the AML Analyzes of the financial institution (FI) that funded the research and the result was compared with other systems in use in that FI. The numbers obtained in the F1-score and *Matthews Correlation Coefficient* metrics allow us to conclude that with the methodology proposed in this thesis it was possible to obtain better results, highlighting that 76% of the profiles confirmed as suspicious were not reported to the regulator at the time of their occurrence, because no system in use at the FI flagged them as suspicious.

Keywords: Anti-Money Laundering; Multi-agent System; Intelligent Agents; Machine Learning; Risk Evaluation

Índice

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

ABC Anti-Branqueamento de Capitais. 1–5, 7, 11, 13, 15, 18, 20, 21, 24, 27, 34, 42–44, 65, 67, 71–73, 77, 81, 89, 90, 92, 93, 95, 97, 103–107, 109

ABP Atualizador da Base de Perfis. 81

ABR Atualizador Base de Regras. 81

AM Aprendizado de Máquina. 30

AML Anti-Money Laundering. 11

AOSE Agent-Oriented Software Engineering. 74, 76, 105

APD Auxiliar do Processo Decisório. 81, 91, 92

AUML Agent Unified Modelling Language. 41, 82

BACEN Banco Central do Brasil. 16, 22, 67

BC Branqueamento de Capitais. 1–5, 7, 8, 10, 12–15, 17, 18, 23, 24, 27, 33, 42–44, 47, 48, 51–53, 63, 64, 67, 71, 74, 86, 89, 95, 96, 99, 103–109

BDI Belief-Desire-Intention. 4, 38, 73, 74, 76, 83, 104, 106

BdP Banco de Portugal. 8, 22

BPstat Banco de Portugal - Estatísticas Online. 19

BRICS Brasil, Rússia, Índia, China e África do Sul. 20

COAF Conselho de Controle de Atividades Financeiras. 8, 10, 12, 16, 22

CPS Capturador de Perfis Suspeitos. 81, 82, 85–87, 91

DOC Documento de Ordem de Crédito. 51, 52, 54, 61

ESOA Engenharia de Software Orientada aos Agentes. 74

EUROSTAT European Commission - statistical office of the European Union. 13

FAIS FinCEN Artificial Intelligence System. 24

FATF Financial Action Task Force. 11, 23, 44

FATFLAT Financial Action Task Force of Latin America. 12

FEBRABAN Federação Brasileira de Bancos. 19

FGCF Fluxo Genérico de Combate a Fraudes. 1, 15, 71

FinCEN Financial Crimes Enforcement Network. 24

FIPA Foundation for Intelligent Physical Agents. 40, 41, 83

FIPA-ACL FIPA Agent Communications Language. 41

FIPA-CCL FIPA Constraint Choice Language. 40

FIPA-KIF FIPA Knowledge Interchange Format. 40, 41

FIPA-RDF FIPA Resource Description Framework. 40

FIPA-SL FIPA Semantic Language. 41

FIPASL FIPA Semantic Language. 40

FIU Financial Intelligence Unit. 12, 13, 16

GAFI Grupo de Ação Financeira Internacional. 11, 12, 14, 23

GAFILAT Grupo de Ação Financeira da América Latina. 11

GAO General Accounting Office. 24

GCS Gerenciador da Captura de Suspeitos. 81, 82, 86, 91

GEC Gerenciador da Evolução do Conhecimento. 82

GRI Global Reporting Initiative. 19

IA Inteligência Artificial. 24, 25, 32, 34, 38

IAD Inteligência Artificial Distribuída. 34

IS Índice de Suspeição. 68, 69, 87, 93, 95, 97, 106, 108

KDD Knowledge Discovery in Databases. 28

KPMG KPMG International Corporate. 72

KQML Knowledge Query and Manipulation Language. 40, 41, 82

KYC Know Your Customer. 11, 14, 16, 23, 32, 43, 44, 84, 95, 106

LD Lavagem de Dinheiro. 7, 8

MAR Margem Adicional de Risco. 65, 66, 86, 87, 91, 92, 99, 100, 105, 106, 108

MCC Matthews Correlation Coefficient. 100–102, 107

MD Mineração de Dados. 28–30, 49, 106

ML Money Laundering. 7

MLP Multilayer Perceptron. 33

MOA Massive Online Analysis. 49, 55

MySQL MySQL Workbench. 46

NTR Número de Transações Reportadas. 14

OIG Office of Inspector General. 24

ONU Organização das Nações Unidas. 11

OTA Office of Technology Assessement. 24

PDT Prometheus Design Tool. 76, 105

PGR Procuradoria Geral da República. 16, 22

PRS Procedural Reasoning Sytem. 82

RBF Radial Basis Function. 32

RDP Resolução Distribuída de Problemas. 34, 39

ROC Receiver Operating Characteristic. 60

RPA Recuperador de Perfis para Análise. 85–87

SAFI Sociedad Anonima Financeira de Inversion. 10

SMA Sistemas Multi-Agente. 34, 35, 39, 40, 73, 106

SSE Sum of Squared Error. 49, 53

SVM Support Vector Machine. 32

TED Transferência Eletrónica Disponível. 51, 52, 54, 55, 61

UIF Unidade de Informação Financeira. 13, 16, 22

UNODC Escritório das Nações Unidas sobre Drogas e Crimes. 1

VRC Variance Ratio Criterion. 49, 53

WEKA Waikato Environment for Knowledge Analysis. 49, 50, 54–56

Capítulo 1

Introdução

1.1 Contexto

Ações de prevenção e combate ao crime de Branqueamento de Capitais (BC) são priorizadas por quase todos os governos do mundo, no mínimo, no mesmo nível das grandes questões globais [85]. BC é, tipicamente, um crime que consiste em tornar licita a origem ilícita de um determinado ganho financeiro. Segundo o Escritório das Nações Unidas sobre Drogas e Crimes (UNODC) a estimativa global anual de branqueamento de capitais é em torno de 2%-5% do Produto Interno Bruto Global, ou US$800 mil milhões - US$2 biliões [121]. Os Estados Unidos continuam a estimar que o crime financeiro doméstico, excluindo a sonegação de impostos, gera aproximadamente US$ 300 mil milhões em recursos passíveis de branqueamento [122]. Este volume financeiro perdido anualmente já é motivo suficiente para o assunto ser tratado com prioridade, no entanto, outro fator leva os governos a priorizarem o combate a este crime: a sua comprovada ligação com outras práticas criminosas como narcotráfico, fraude, corrupção, sequestro, terrorismo, contrabando de armas, entre outros [112].

As autoridades financeiras dos países, normalmente Bancos Centrais, são responsáveis pelo controle e definição de normas regulatórias de Anti-Branqueamento de Capitais (ABC), exigindo das instituições financeiras a implementação de procedimentos que apliquem referidos normativos. Vale ressaltar que, conforme será mostrado no Processo de Deteção, Análise e Comunicação de Indícios (subsecção 2.2 - pág. 14), o processo de ABC nas instituições financeiras é uma instância de um Fluxo Genérico de Combate a Fraudes (FGCF), estando, em sua maioria, semiautomatizados na fase de sinalização de transações suspeitas de BC. Atualmente estes processos utilizam médias, desvios-padrões e regras fixas pré-estabelecidas, regras estas geralmente baseadas em observações empíricas ou na experiência humana dos Analistas de ABC.

1.2 Motivação

A capacidade dos criminosos de se autofinanciar, em função do lucro fácil, permite-lhes adotar procedimentos cada vez mais sofisticados visando burlar as regras de combate ao crime. Esta situação decorre, principalmente, devido ao fato do processo de ABC apresentar entraves relevantes, tais como:

- a publicação das regras de combate e prevenção não tem caráter sigiloso. Desta forma, torna-se um conhecimento de fácil acesso, levando ao desenvolvimento, por parte dos infratores, de estratégias de dissimulação das operações;
- o processo de ABC não apresenta uma etapa de retroalimentação que possibilite, objetivamente, associar as transações sinalizadas como suspeitas com a real comprovação de atividades ilícitas. Esta ausência dificulta a otimização da fase de sinalização de transações suspeitas que, aliado ao aumento constante no volume de transações efetuadas diariamente, impõe aos Analistas de ABC uma sobrecarga de transações a serem analisadas;
- abordando ainda o processo de ABC, ele pode ser resumido com as fases de identificação, análise e reporte. Considerando a ineficiência na identificação dos casos e a ausência de retroalimentação, muitos casos não são analisados e, portanto, não são reportados, impossibilitando a criação de novas regras baseadas na experiência adquirida, tanto por parte da instituição financeira como dos órgãos reguladores. Esta situação pode ser caracterizada como um risco sistémico;
- a geração de novas regras para prevenção e combate dependem da disponibilidade de tempo dos Analistas de ABC, considerando que sua atividade principal é a análise das transações suspeitas;
- a semiautomatização do processo é incapaz de reduzir, significativamente, a quantidade de casos suspeitos destinados a cada Analista de ABC;
- desde o surgimento do processo de ABC as análises são realizadas por especialistas humanos que detêm o conhecimento das normas vigentes e das principais táticas utilizadas para o branqueamento, apresentando relato de cada caso para tomada de decisão quanto ao reporte. Até o momento, nenhum sistema ou método automatizado proposto ou em uso apresentou eficiência que fosse revertida em confiança suficiente que resultasse na dispensa a ação humana na análise. Importante salientar que a análise de um caso de BC ainda utiliza muita subjetividade para sua conclusão.

A consequência imediata desses entraves é tornar a atividade de prevenção e combate ao crime de BC pouco eficiente, transcorrendo num lapso de tempo cada vez maior e já há muito inadequado [28].

Uma possível explicação para esta falta de eficiência do processo é que as instituições financeiras se encontram diante de um verdadeiro círculo vicioso. A dinâmica do mercado e o avanço tecnológico do setor bancário provoca um aumento vertiginoso no volume de transações realizadas e, consequentemente, no volume de transações suspeitas

a serem analisadas. A solução imediata é a alocação de mais Analistas ao processo, sendo que os entraves citados mantêm a ineficiência do processo e cada vez mais humanos são necessários. Sendo que, por alegadas questões de custo, esta alocação não acontece na forma adequada.

A situação descrita levou a que os trabalhos realizados por pesquisadores, utilizando técnicas de Inteligência Artificial, tivessem um início efetivo de utilização nas instituições financeiras. Houve a perceção da capacidade dessas técnicas de analisar um volume maior de dados e de combiná-los com novas fontes de informação, permitindo detetar anomalias ou padrões que caso contrário, passariam despercebidos [57].

1.3 Questões de Pesquisa e Hipótese Principal

As publicações disponíveis descrevendo estudos sobre a utilização de técnicas computacionais no combate ao crime de Branqueamento de Capitais (BC) concentram-se, basicamente, em duas linhas: 1) proposta de novos algoritmos ou melhoria em algoritmos existentes visando a sinalização de casos suspeitos de BC [132, 119, 80, 79, 69]; 2) proposta de sistemas que buscam melhorar e agilizar o processo de sinalização de transações suspeitas [65, 131, 64, 103, 117].

Com relação aos algoritmos, estes estudos carecem de validação da eficácia da proposta perante uma quantidade de dados mais próxima da realidade diária de uma instituição financeira. Os sistemas propostos, mesmo os baseados em multi-agentes incorporam pouco da experiência dos Analistas de ABC [64, 131, 103], sendo poucas as abordagens baseadas em risco.

Diante da motivação apresentada anteriormente e desta situação acima descrita, as seguintes questões de pesquisa podem ser elaboradas:

a. Os algoritmos consagrados de aprendizagem indutiva não-supervisionada e a aplicação de regras por meio de agentes inteligentes são técnicas suficientes para detetar com eficácia casos suspeitos de BC, em grandes volumes de dados? se não, como torná-las?
b. Agentes inteligentes podem ser desenhados para incorporar parte da experiência dos Analistas de ABC e auxiliá-los no processo de tomada de decisão sobre os casos suspeitos de BC?
c. O processo de ABC pode ser desenhado de forma genérica, permitindo a definição de um sistema multi-agente que identifique e decida sobre casos suspeitos de BC, apresentando melhorias face ao estado atual?

Para solucionar estas questões foi assumido como hipótese que “agentes inteligentes que incorporem técnicas de *data mining*, aprendizagem indutiva e autónoma, identificam com eficácia casos suspeitos de BC e apresentam índice satisfatório de decisões autónomas na análise dos casos suspeitos”.

1.4 Objetivos

O objetivo principal desta tese é

> "a formalização de um novo fluxo de tomada de decisão, combinando mineração de dados e agentes inteligentes, que possa ser implementado num sistema de suporte ao processo de Anti-Branqueamento de Capitais (ABC)."

Os objetivos relacionados a seguir são classificados como específicos:

1. Ampliar e sistematizar o conhecimento sobre os processos inerentes ao ABC;
2. Mapear o processo de tomada de decisão praticado por um Analista de ABC no exercício da tarefa de análise das transações sinalizadas como suspeitas;
3. Conceber um modelo de agentes inteligentes que permita representar o conhecimento sistematizado no processo de ABC;
4. Definir e implementar um sistema computacional que atenda os seguintes princípios básicos:
   4.1. Tornar mais eficaz e inteligente a identificação de transações suspeitas;
   4.2. Dotar de moderada autonomia a tomada de decisão de parte do processo de análise das transações suspeitas;
   4.3. Auxiliar e aprender com o Analista de ABC na tomada de decisão sobre os casos mais complexos de transações suspeitas;
   4.4. Criar subprocesso de sugestão de novas regras baseadas em novos critérios.

1.5 Contribuições

Ao atingir os objetivos elencados também terá sido possível oferecer contribuições significativas para a área do conhecimento em informática e, consequentemente, para a área a instituição financeira fornecedora dos dados e patrocinadora do estudo. Estas contribuições podem ser assim relacionadas:

- Mudança de paradigma na abordagem genérica de combate ao crime de Branqueamento de Capitais (BC);
- Solução baseada em agentes, num processo de cooperação de análise, para tomada de decisão e mitigação do risco sistémico identificado no processo de Anti-Branqueamento de Capitais (ABC);
- Metodologia de utilização de técnicas de *data mining* aplicável a grande volume de dados visando agrupamento de clientes induzido sob uma perspetiva de grupos de risco;
- Modelo de integração entre a metodologia de *data mining* e a arquitetura de sistema baseado em agentes Belief-Desire-Intention (BDI);

1.6 Estrutura deste Documento

O desenvolvimento da proposta apresentada nesta introdução está descrito neste documento, dividido na seguintes estrutura:

O *Capítulo 2* apresenta conceitos sobre a natureza e tipificação do crime de BC, características sobre o combate a este crime e os gargalos identificados nesta atuação. Expõe as principais estatísticas deste crime no Brasil e em Portugal, permitindo observar o desafio da volumetria envolvida. Descreve o erro sistémico existente no processo de ABC, identificado neste trabalho e descreve as primeiras iniciativas visando a automatização deste processo.

As principais técnicas e abordagens utilizadas no apoio ao ABC estão descritas no *Capítulo 3*, com ênfase na mineração de dados, aprendizado de máquina e sistemas baseados em agentes inteligentes.

O *Capítulo 4* descreve e quantifica os dados utilizados no trabalho, realizando um histórico das principais definições visando o atingimento dos objetivos definidos. Também relaciona as ferramentas e técnicas utilizadas no desenvolvimento da solução.

Detalhes sobre a modelagem do sistema são apresentados no *Capítulo 5*, que também apresenta exemplos dos códigos utilizados.

O *Capítulo 6* apresenta e analisa os resultados obtidos.

As conclusões e sugestões para trabalhos futuros estão no *Capítulo 7*.

Além desses capítulos o documento também contém apêndices onde estão os diagramas gerados pela metodologia de desenvolvimento de sistema utilizada no projeto e detalhes das bases de dados utilizada no produto a ser implementado.

Capítulo 2

Branqueamento de Capitais, Conceitos, Combate e Gargalos

A capacidade de adaptação do *modus operandi* dos fraudadores e a falta de informação sistematizada que associe as transações suspeitas com a comprovação do crime, são fatores que dificultam demasiado a automatização do processo de prevenção e combate ao crime de Branqueamento de Capitais (BC). É possível que esta seja a explicação para as raras soluções inovadoras tornadas de conhecimento público, no âmbito do Anti-Branqueamento de Capitais (ABC). O objetivo deste capítulo é apresentar um panorama sobre ações mundiais visando o combate ao crime de BC, mostrar o estágio atual em Portugal e no Brasil, ressaltando os grandes problemas enfrentados por quem tem a obrigação de combater este crime.

2.1 Branqueamento de Capitais - Introdução e Conceitos

Quer seja na indústria financeira, nos órgãos centrais de controle e regulamentação da atividade financeira ou nas ações de governo de todas as nações do mundo, a busca pelo aprimoramento do processo de prevenção e combate a crimes que podem provocar perda de capitais está no topo da relação de prioridades. Lavagem de Dinheiro (LD), Branqueamento de Capitais (BC) ou *Money Laundering (ML)*, é um dos crimes que tem por objetivo a legalização de capitais obtidos de forma ilícita, na maioria dos casos com propósito de financiamento de atividades ilegais. As organizações e governos têm empreendido esforços, quer isoladamente ou em cooperação, visando o mapeamento do *modus operandi* desses criminosos.

A já comprovada associação entre este crime e os crimes de narcotráfico, corrupção, sequestro, terrorismo, contrabando de armas, entre outros, justifica esta preocupação mundial com o tema [112]. As nações têm promovido alterações e inovações nas suas legislações e nos seus procedimentos institucionais, sempre buscando um melhor combate às cada vez mais diversas e audaciosas modalidades desse crime [85].

É chamado de "dinheiro sujo" (*dirty money*) aquele que resulta do lucro com a prática

de crimes, como os citados anteriormente, cujos autores, geralmente, pertencem a uma organização criminosa [15].

Para Mendroni a prática de legalizar recursos obtidos de forma ilícita, remonta ao século XVII com os piratas [87]. Contudo, a origem do termo *money laundering* remonta ao tempo das quadrilhas de *gangster* americanos nas décadas de 1920 e 1930, que utilizavam lavandarias de roupas e lavadores de autos para dissimular a origem criminosa dos seus recursos financeiros [110].

Castellar esclarece sobre a origem do termo quando afirma que [30, p. 81]:

> "A expressão lavagem de dinheiro originou-se, historicamente, no costume das máfias norte americanas, na segunda década do século 20, de usar lavandarias para ocultar a procedência ilegal de seu dinheiro. Deve-se observar que em muitos países, inclusive Portugal, em vez de 'lavagem de dinheiro' é usado o termo 'branqueamento de dinheiro'. Internacionalmente, a expressão 'money laundering' é utilizada para designar esta atividade. "

Na atualidade o termo oficialmente utilizado em Portugal é Branqueamento de Capitais (BC), o qual será adotado, de agora em diante, neste trabalho, exceto nas citações de autores brasileiros onde o termo Lavagem de Dinheiro (LD) é utilizado.

Na tipificação dos crimes relacionados ao BC, aqueles que provocam ganho de capital e que precisam ser "branqueados" para retornar ao mercado sem que sua origem seja descoberta, são chamados crimes antecedentes [33].

Segundo o Conselho de Controle de Atividades Financeiras (COAF) o BC caracteriza-se por um conjunto de operações comerciais ou financeiras que buscam a incorporação na economia de cada país, de modo transitório ou permanente, de recursos, bens e valores de origem ilícita [34].

O Banco de Portugal (BdP) define branqueamento de capitais como "o processo pelo qual os autores de algumas atividades criminosas encobrem a origem dos bens e rendimentos (vantagens) obtidos ilicitamente, transformando a liquidez proveniente dessas atividades em capitais reutilizáveis legalmente, por dissimulação da origem ou do verdadeiro proprietário dos fundos" [12].

Além da similaridade nestas definições, universalmente há concordância quanto as ações criminosas se desenvolverem por meio de um processo dinâmico constituído, teoricamente, de três fases independentes (colocação, ocultação e integração) que, com frequência, ocorrem simultaneamente (Figura 2.1).

Estas fases do processo são assim descritas pelo COAF na cartilha Lavagem de Dinheiro: Um problema Mundial [34, p. 4]:

> "Colocação – A primeira etapa do processo é a colocação do dinheiro no sistema económico. Objetivando ocultar sua origem, o criminoso procura movimentar o dinheiro em países com regras mais permissivas e naqueles que possuem um sistema financeiro liberal. A colocação se efetua por meio de depósitos, compra de instrumentos negociáveis ou compra de bens. Para

dificultar a identificação da procedência do dinheiro, os criminosos aplicam técnicas sofisticadas e cada vez mais dinâmicas, tais como o fracionamento dos valores que transitam pelo sistema financeiro e a utilização de estabelecimentos comerciais que usualmente trabalham com dinheiro em espécie.

Ocultação – A segunda etapa do processo consiste em dificultar o rastreamento contábil dos recursos ilícitos. O objetivo é quebrar a cadeia de evidências ante a possibilidade da realização de investigações sobre a origem do dinheiro. Os criminosos buscam movimentá-lo de forma eletrónica, transferindo os ativos para contas anónimas – preferencialmente, em países amparados por lei de sigilo bancário – ou realizando depósitos em contas 'fantasmas'.

Integração – nesta última etapa, os ativos são incorporados formalmente ao sistema económico. As organizações criminosas buscam investir em empreendimentos que facilitem suas atividades – podendo tais sociedades prestarem serviços entre si. Uma vez formada a cadeia, torna-se cada vez mais fácil legitimar o dinheiro ilegal."

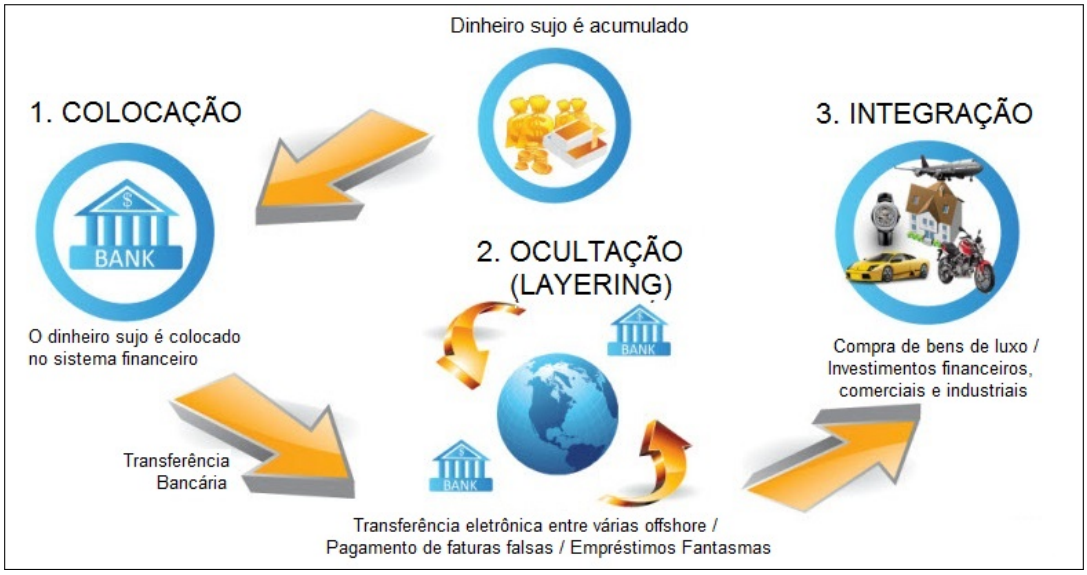


Figura 2.1: Ciclo do processo de branqueamento de capitais[1]

Canas [28] acrescenta que na fase de colocação a introdução de bens e produtos também pode ocorrer em pontos do circuito financeiro tais como: casas de câmbio, casinos e instituições de investimento. Na ocultação, que o autor chama de camuflagem, ele salienta a criação de "camadas" (*layering*) envolvendo transformação em várias moedas e sucessivas transferências, tornando quase impossível identificar a origem. Na integração

[1]Figura adaptada de http://paulrenner.com/C6-Intelligence/paul-renner-c6-kycmap-what-is-money-laundering.html

o autor inclui a compra de veículos de comunicação social na relação de bens buscados na legitimação do dinheiro devidamente branqueado.

Por esta descrição é possível perceber que para disfarçar os lucros ilícitos e não comprometer os envolvidos, o BC impõe dinamismo no processo quando requer: primeiro, o distanciamento dos fundos de sua origem, evitando uma associação direta deles com o crime; segundo, o disfarce de suas várias movimentações para dificultar o rastreamento desses recursos; e terceiro, a disponibilização do dinheiro novamente para os criminosos depois de ter sido suficientemente movimentado no ciclo de lavagem e poder ser considerado "limpo".

Ainda na cartilha sobre o processo de BC o COAF utiliza o didático caso Franklin Jurado, conduzido pela justiça americana entre 1990 e 1996, para ilustrar um ciclo clássico [34, p. 5]:

> "Economista colombiano formado em Harvard, Jurado coordenou a lavagem de cerca de US$ 36 milhões em lucros obtidos por José Santacruz Londono com o comércio ilegal de drogas.
>
> O depósito inicial - o estágio mais arriscado, pois o dinheiro ainda está próximo de suas origens - foi feito no Panamá. Durante um período de três anos, Jurado transferiu dólares de bancos panamenses para mais de 100 contas diferentes em 68 bancos de nove países, mantendo os saldos abaixo de US$ 10 mil para evitar investigações.
>
> Os fundos foram novamente transferidos, dessa vez para contas na Europa, de maneira a obscurecer a nacionalidade dos conta-correntistas originais, e, então, transferidos para empresas de fachada. Finalmente, os fundos votaram à Colômbia por meio de investimentos feitos por companhias europeias em negócios legítimos, como restaurantes, construtoras e laboratórios farmacêuticos, que não levantariam suspeitas.
>
> O esquema foi interrompido com a falência de um banco em Mónaco, quando várias contas ligadas a Jurado foram expostas. Fortalecida por leis antilavagem, a polícia começou a investigar o caso e Jurado foi preso."

Para Satula [110] a maioria dos autores que aborda o fenómeno acaba por caracterizalo, quanto a forma de execução, em: internacionalização da atividade; grande volume de transações; especialização das organizações envolvidas; e diversidade de técnicas.

A internacionalização talvez explique porque a quase totalidade dos casos de BC, tornados públicos, mostram que para o sucesso do processo é necessária a existência dos denominados paraísos fiscais, com facilitação das instituições financeiras, comerciais e com aval ou conivência do governo local. Nesses paraísos fiscais são utilizadas empresas genericamente denominadas *off-shore*, que, no entanto, adotam denominações diferentes em outras localidades, tais como *Exempt Companies* nas Ilhas Virgens Britânicas ou *Sociedad Anonima Financeira de Inversion (SAFI)*, no Uruguai.

Conclusivamente, Canas atesta que "apesar da diversidade dos métodos e das técnicas de branqueamento, a maior parte das operações quase sempre passarão, em algum momento, pelo sistema financeiro" [28, p. 81].

2.1.1 Evolução das Ações de Anti-Branqueamento de Capitais

É possível indicar o final da década de 1980 como o momento em que a comunidade internacional desencadeou esforços no sentido de criar leis, organizações e mecanismos de controle para o combate ao crime de branqueamento de capitais. O objetivo inicial era harmonizar, no que fosse possível, as legislações internas de cada País, "para impedir práticas contrárias a ordem pública, aos bons costumes, que ofendessem os postulados do *fair play* e de justiça substantiva" [110, p. 48].

Como consequência, surgem no final de 1988 dois documentos internacionais pioneiros na definição de regras para o combate ao BC: A Declaração de Basileia e a Convenção da Organização das Nações Unidas (ONU).

A chamada Declaração de Basileia é o documento *Prevention of Criminal Use of the Banking System for the Purpose of Money Laundering* [16], que apresenta conceitos que orientam: a criação da política de Conheça seu Cliente (*Know Your Customer (KYC)*); a manutenção da conformidade com as leis e as boas práticas de mercado; a formação de um sistema interno de controle nas instituições financeiras; bem como o treinamento dos funcionários em controle interno. Estas orientações resultaram em procedimentos que pela primeira vez trataram do tema Anti-Branqueamento de Capitais (ABC) ou *Anti-Money Laundering (AML)*. Estes conceitos foram reforçados no depoimento apresentado pelo vice-presidente do Conselho do Federal Reserve, Roger W. Ferguson, Jr., em 18/06/2003, perante o Comité do Senado dos EUA para Assuntos Bancários, Habitação e Urbanos, publicado no boletim do *The Federal Reserve Board* dos EUA [56].

A Convenção das Nações Unidas Contra o Tráfico Ilícito de Estupefacientes e de Substâncias Psicotrópicas, chamada Convenção de Viena [93], apesar do título, estava mais voltada para o tema ABC. Um dos motivos prováveis foi porque em convenções anteriores a comunidade internacional havia percebido que os traficantes estavam modernizando e tornando mais complexo seus meios de financiamento, o que exigia estratégias específicas para o combate à este crime organizado, reforçando a atividade de ABC. A Convenção então oficializou o compromisso internacional para a criminalização da prática descrita.

O Grupo de Ação Financeira Internacional (GAFI) ou *Financial Action Task Force (FATF)* é um órgão intergovernamental criado em julho de 1989 pela Reunião de Cúpula do G-7 realizada em Paris[2], França. O GAFI tem divulgado recomendações que são reconhecidas como padrão internacional para o ABC, ao combate ao financiamento do terrorismo e a proliferação de armas de destruição em massa.

Outros órgãos, regionalizados, ligados ao GAFI foram criados, como o Grupo de

[2]http://www.fatf-gafi.org/about/

Ação Financeira da América Latina (GAFILAT) ou *Financial Action Task Force of Latin America (FATFLAT)*, criado em novembro de 2006, do qual o Brasil faz parte[3], além de também integrar o GAFI desde 2000. Na Península Ibérica não existe nenhum órgão específico, por isso, Portugal é exclusivamente membro do GAFI desde 1990.

O Grupo de Egmont, criado em 1995, resultante de reunião ocorrida no Palácio de Egmont-Aremberg, em Bruxelas[4], consiste de uma rede internacional de cooperação, com sede em Toronto no Canadá, formada pelas *Financial Intelligence Unit (FIU)* dos atuais 164 países membros. O principal compromisso assumido por estes países foi:

> "Cada País subscritor do acordo internacional comprometeu-se a criar sua própria 'Unidade de Inteligência Financeira', que tem a função de receber e concentrar as informações a respeito das operações suspeitas de encobrir atividade de lavagem de dinheiro e repassá-las aos órgãos de persecução com atribuição para a investigação e processamento criminal – por excelência –, polícias e Ministérios Públicos." [87, p. 483]

As recomendações contidas nos documentos citados, bem como aquelas emitidas e frequentemente atualizadas pelos órgãos reguladores criados, geram um grande volume de informações a serem tratadas neste processo. Volume sempre crescente de normas, recomendações e de dados a serem analisados e compartilhados.

2.1.2 Base Legal e Estatísticas no Brasil e em Portugal

No Brasil a Lei 9.613, de 3/3/1998 [81], tipificou o crime de Branqueamento de Capitais (BC) como aquele em que se oculta ou dissimula a natureza, a origem, a localização, a disposição, a movimentação ou a propriedade de bens, direitos e valores provenientes, direta ou indiretamente, de determinados crime antecedentes [33].

O órgão especializado para averiguar a prática de operações de BC, nos moldes de uma FIU, é o Conselho de Controle de Atividades Financeiras (COAF), que, criado pela citada Lei 9.613/1998, visa a implementação de políticas nacionais voltadas ao combate ao BC.

Em seu último Relatório de Atividades o COAF informa suas atividades e principais realizações, constando também o total de comunicações recebidas e a quantidade de Relatórios de Inteligência Financeira produzidos no período, conforme as Figuras 2.2 e 2.3 [35].

Em Portugal o crime de branqueamento de capitais foi tipificado pela Lei 25/2008 de 5 de junho [82], que estabelece obrigações para as entidades financeiras, nomeadamente instituições de crédito, empresas de investimento e outras sociedades financeiras, empresas seguradoras que exerçam atividades no ramo vida e sociedades gestoras de fundos de pensões. A lei designa para as entidades financeiras, dentre outros, o chamado dever

[3] http://www.fatf-gafi.org/pages/members/financialactiontaskforceoflatinamericagafilat.html
[4] http://www.egmontgroup.org/

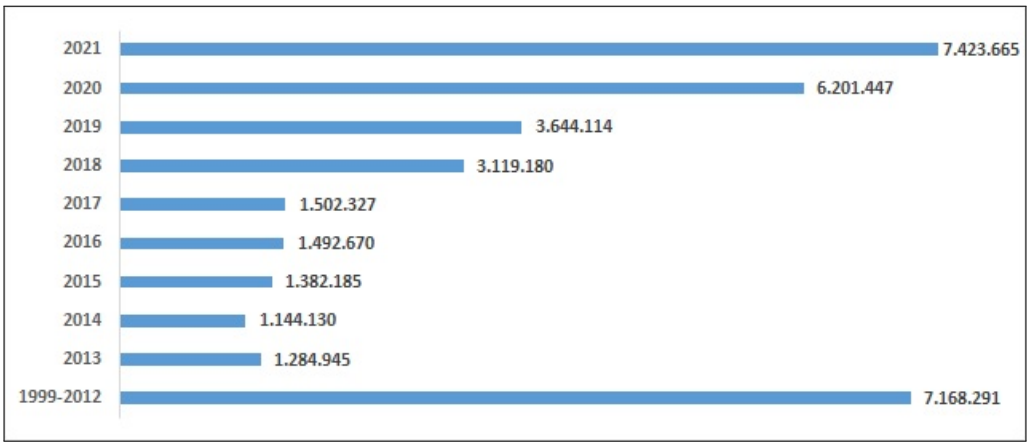


Figura 2.2: Comunicações recebidas pelo COAF (Brasil)

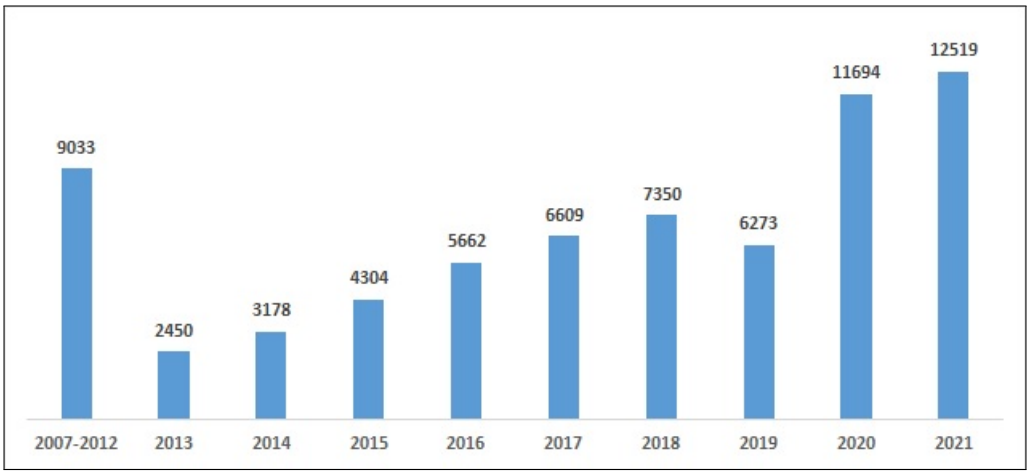


Figura 2.3: Relatórios de Inteligência Financeira emitidos pelo COAF (Brasil)

de comunicação: se uma operação suspeita não puder deixar de se realizar, a entidade executa-a e deve fornecer de imediato à autoridade judiciária, no caso o Procurador Geral da República, todas as informações respetivas, de modo a permitir que este execute sua atividade de investigação [28, 110].

O órgão responsável por averiguar a prática de operações de BC, nos moldes de uma FIU, é a Unidade de Informação Financeira (UIF) subordinada aos Serviços da Direção Nacional, órgão da Polícia Judiciária, criada pelo Decreto-Lei nº 304/2002, de 13 de dezembro [42]. As suas competências estão descritas no Decreto-Lei nº 42/2009, de 12 de fevereiro [43], e na própria lei de prevenção e combate ao BC [82].

Um dos objetivos da Comissão Europeia é facilitar e dar suporte ao ABC, além de difundir informações e estatísticas sobre este crime, no entanto, a complexidade dessa tarefa fez com que somente em 2010 fosse realizado pelo *European Commission - statistical office of the European Union (EUROSTAT)*, situado em Luxemburgo, o primeiro relatório com dados estatísticos sobre Branqueamento de Capitais na Europa, denominado *Money Laundering in Europe* [50].

Como comentado anteriormente, as 40 Recomendações do Grupo de Ação Financeira Internacional (GAFI) são reconhecidas como um padrão global contra o BC e o financiamento ao terrorismo. O GAFI publica regularmente o relatório *Anti-money laundering and counter-terrorist financing measures - Mutual Evaluation Report* com informações e estatísticas sobre cada país membro. O último relatório com dados de Portugal foi publicado em 2017 [53]. Baseado neste relatório, a Figura 2.4 mostra a série de Número de Transações Reportadas (NTR).

Figura 2.4: Comunicações recebidas pelo FATF (Portugal)

Para o mesmo relatório Portugal reportou a quantidade de processos instaurados e as condenações realizadas com base no crime de BC, conforme mostra a Tabela 2.1.

Tabela 2.1: Processos e Condenações informadas ao FATF (Portugal)

Branqueamento de Capitais	**2010**	**2012**	**2013**	**2014**	**2015**	**2016**
NTR	2.281	3.047	2.776	3.910	5.047	5.368
Processos Instaurados	80	60	92	77	115	123
Condenações	13	19	36	19	14	-

2.2 Processo de Deteção, Análise e Comunicação de Indícios

O Comité de Basileia, conforme já comentado, definiu em [16] a política de Conheça seu Cliente (*Know Your Customer (KYC)*), como boa prática de combate a fraudes. Os princípios do KYC foram reforçados em [17], documento que detalha os procedimentos a serem seguidos. Tendo por base estes documentos, bem como o *benchmark* da empresa Hewlett-Packard (HP) publicado pelo *The Institute of Internal Auditors* no artigo [78],

é possível realizar um mapeamento da prática genérica de fraude ou burla na atividade económica, cuja operacionalização resulte na execução de uma operação formal, direta ou indiretamente informatizada. Este mapeamento mostra a existência de elementos comuns na atividade criminosa, independente do setor onde é praticado. Consequentemente, o combate a este crime também resulta em procedimentos similares. O Fluxo Genérico de Combate a Fraudes (FGCF), mostrado na Figura 2.5, é uma representação gráfica deste citado mapeamento. Importante observar que apesar do Branqueamento de Capitais (BC) ser um crime adjacente, ou seja, "para haver Branqueamento teria de haver um crime anterior que proporcionasse ilicitamente ao seu autor proventos que posteriormente ele, ou outrem, pretendesse camuflar" [28], na sua essência é possível afirmar que ele é uma instância de um processo genérico de fraude ou burla. Dessa forma, a atividade de Anti-Branqueamento de Capitais (ABC) acaba sendo o fluxo genérico mostrado aplicado ao setor financeiro.

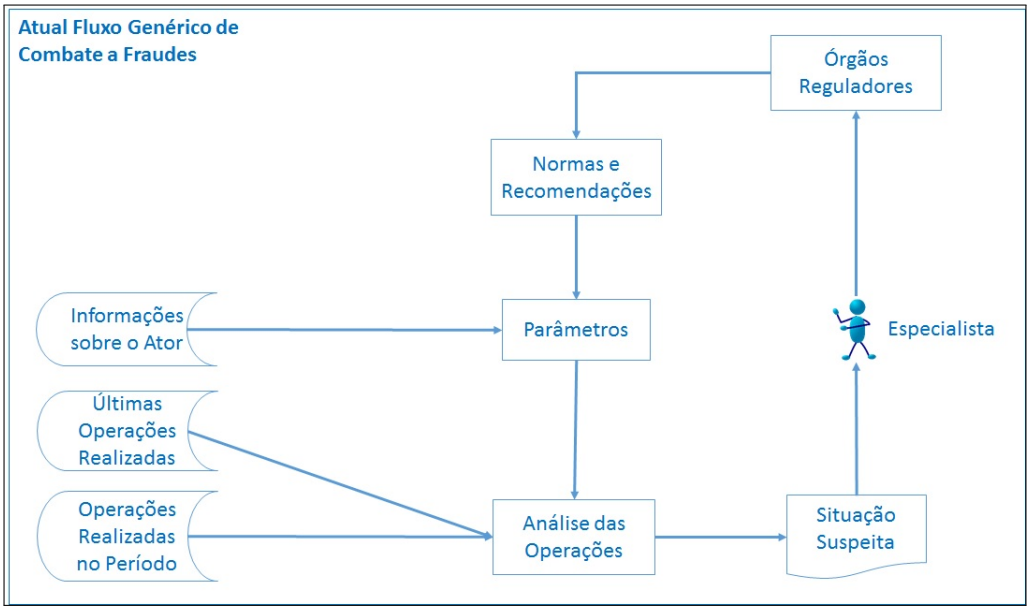


Figura 2.5: Fluxo Genérico de Combate a Fraudes (Atual)

Um ponto em comum entre as leis e normas, emanadas dos órgãos de controle citados anteriormente, diz respeito aos deveres das entidades sujeitas à esta base legal. Dessa forma, têm responsabilidade específica, todas as pessoas físicas e jurídicas que exerçam atividades, principal ou acessória, relacionadas a: captação, intermediação ou aplicação de recursos financeiros; compra e venda de moeda estrangeira, ouro, ativo financeiro ou instrumento cambial; negociação ou administração de títulos ou valores mobiliários; gestão ou comercialização de fundos de pensão ou de capitais de risco [81, 82].

Essas responsabilidades específicas estão claramente definidas na legislação portuguesa, Lei 25/2008 [82], enunciados no artigo 6º, como deveres gerais: de identificação, de diligência, de recusa, de conservação, de exame, de comunicação, de abstenção, de colaboração, de segredo, de controle e de formação.

Dentre as entidades sujeitas a esse arcaboiço legal estão as instituições financeiras, em especial os Bancos, que, em todo o mundo, tiveram de redesenhar e até criar processos visando atender esses preceitos. Alguns dos deveres citados tiveram maior relevância na reestruturação ocorrida nos Bancos, tais como:

- Exigir documentos de identificação e verificar a identidade de clientes e representantes (*KYC*);
- Exercer acompanhamento contínuo dos clientes e de seus negócios, mantendo sempre atualizado e disponível seu perfil de risco (diligência);
- Manter registo físico e digital do cumprimento dos deveres e das transações realizadas (conservação);
- Efetuar análise sobre atividade e transações cujos elementos acusem incompatibilidade e possam ser relacionados com o branqueamento de capitais (exame); e
- Informar de imediato o órgão competente (COAF no caso do Brasil, UIF e Procuradoria Geral da República (PGR) no caso de Portugal) quando do conhecimento, suspeita ou existência de fatos indiciários da prática de crime de branqueamento de capitais.

O sistema financeiro é constituído por um conjunto de instituições, mercados e recursos de um determinado país, voltados para a viabilização de transações com promessas de pagamento a ser realizada no futuro, por pessoas ou instituições, que se tornam assim devedores; e que são aceitas por outras pessoas ou instituições, que se tornam desta forma credores dos primeiros [29].

Dentro deste já complexo sistema foi então criado, em cada país, um processo para abrigar os órgãos de controle citados anteriormente e cumprindo as exigências e recomendações emanadas dos acordos internacionais e das legislações locais. Um esboço genérico desse processo está apresentado na Figura 2.6.

Considerando que tanto Brasil quanto Portugal têm órgãos de acompanhamento e controle nos moldes de uma FIU (COAF e UIF) e que têm, também, os respetivos Bancos Centrais (Banco Central do Brasil e Banco de Portugal) como órgãos reguladores e legisladores, o processo do Sistema Financeiro de ambos apresentam poucas diferenças. Cabe ressaltar que o COAF é subordinado ao BACEN. As instituições financeiras, nomeadamente os bancos, têm por dever comunicar todas as situações suspeitas e isso é feito remetendo algumas informações para o COAF e outras para o Banco Central do Brasil (BACEN), no caso do Brasil, e para a UIF e PGR, no caso de Portugal. Em ambos os casos essas instituições de acompanhamento e controle também reportam os casos para os respetivos Bancos Centrais e outros órgãos competentes, conforme o caso. Com base no *modus operandi* e no aprendizado obtido com os casos reportados, além da adaptação das recomendações internacionais, os Bancos Centrais formulam normas e recomendações que devem ser seguidas pelas instituições financeiras. Também as FIU podem emitir recomendações visando o aprimoramento do processo.

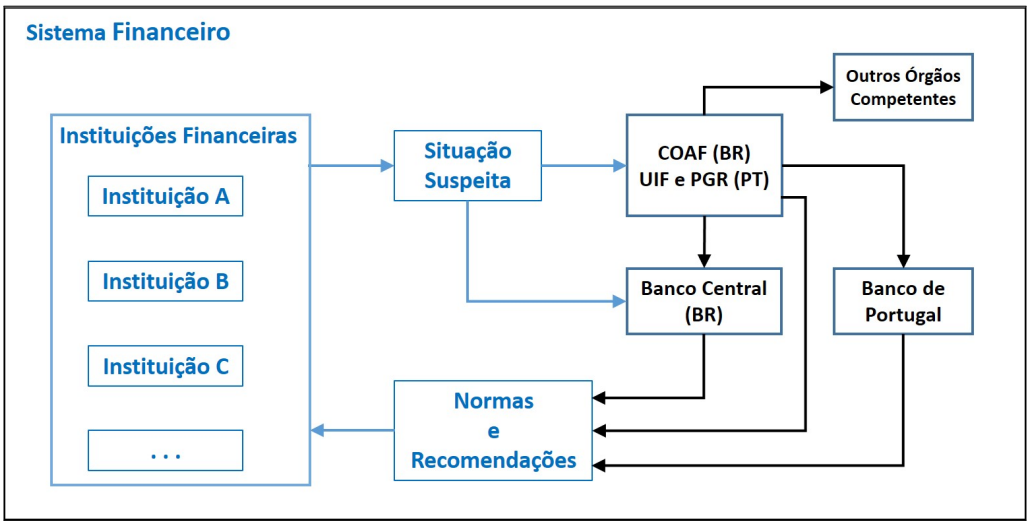


Figura 2.6: Sistema Financeiro: Processo Genérico de ABC

É possível perceber que em função do conjunto de leis de cada país, bem como as especificidades de cada mercado financeiro, os processos acabem tornando-se específicos para cada país e até para cada instituição financeira. No entanto, detalhando um pouco mais o processo mostrado na Figura 2.6, um possível fluxo genérico para o processo interno de uma instituição financeira, é sugerido na Figura 2.7.

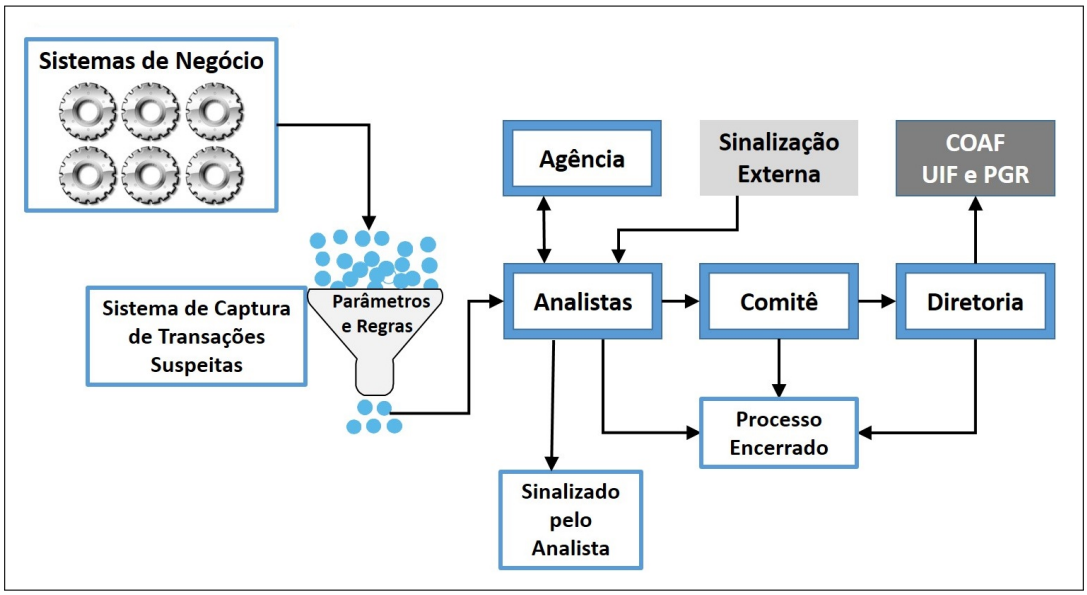


Figura 2.7: Instituição Financeira: Processo Genérico de ABC

As transações são geradas pelos sistemas funcionais de negócio e algumas são capturadas como suspeitas por sistemas especializados que verificam limites predefinidos. Em seguida essas transações são submetidas a um conjunto de regras e parâmetros, que congregam as diretrizes legais para a prevenção e combate ao crime de BC com as orientações produzidas pela Área de Risco da instituição. Normalmente, este filtro é genérico

e conservador, resultando num grande conjunto de transações que devem ser analisadas por um especialista humano.

O especialista, normalmente conhecido como Analista de *Compliance* ou Analista de ABC, utiliza seus conhecimentos e experiências baseadas na norma vigente para decidir entre: pedir mais informações à agência que atende o cliente que executou a transação; sinalizar o cliente para um acompanhamento mais rigoroso; encerrar a análise daquela transação ou considerar procedente e encaminhar para o Comité recomendando o despacho de transação suspeita para a Diretoria. A decisão final é da Diretoria da instituição que, entendendo procedente, comunica a autoridade competente, ou, caso contrário, encerra o processo.

Frequentemente o Comité é composto por executivos de várias áreas da empresa e podem deliberar por não acatar a recomendação do Analista, encerrando o processo, ou encaminhar para a Diretoria da instituição.

É possível perceber um elevado nível de subjetividade e grande lapso de tempo entre a identificação da suspeita e a decisão final, questões que serão tratadas a seguir.

2.2.1 Complexidade e Gargalos do Processo de ABC

Uma observação superficial no fluxo mostrado na Figura 2.7 pode passar ao observador a impressão de simplicidade, pressupondo que basta identificar indícios ou suspeitas e comunicar às autoridades competentes.

No entanto, exatamente esses dois pontos salientados (identificar e comunicar) tornam o processo complexo e crítico. A complexidade está em torno da tarefa de identificar e analisar as transações suspeitas; e a criticidade reside em executar estas tarefas em tempo hábil, mesmo que não consiga impedir a primeira ocorrência, mas que permita evitar a repetição e propagação do crime.

As instituições financeiras, particularmente os bancos, enfrentam dificuldades com a volumetria das transações que devem ser analisadas diariamente, numa diversidade cada vez maior de canais de atendimento, confrontando-as com o grande arcabouço legal existente, impedindo a celeridade do processo.

Outro fator impactante é a inexistência de mecanismos tecnológicos eficientes que agilizem a identificação e análise dos indícios do crime de Branqueamento de Capitais (BC), fazendo com que esta fase ocorra com indesejável desfasagem de tempo, podendo chegar a meses[5]. Este é um prazo inaceitável, pois permite que o crime seja cometido e aqueles que o praticam possam se evadir, contudo, com o atual modelo de análise, um prazo menor pode ser inviável.

O problema envolvendo o tempo adequado para análise de uma transação suspeita foi alertado por Canas ao afirmar que:

> "um quadro legal que pressione demasiado o sistema bancário, ou qualquer

[5]O COAF estabelece prazo máximo de 90 dias para comunicação de uma situação suspeita, a partir do momento que a transação é realizada

> outro sistema ou conjunto de entidades, a realizar um número apreciável e crescente de comunicações prévias pode conduzir a uma situação em que os meios de investigação criminal não conseguem acompanhar e analisar com rigor mais que uma ínfima parte dessas comunicações, inutilizando e descredibilizando o sistema de prevenção" [28, p. 190].

2.2.2 Obstáculo da Volumetria

A Federação Brasileira de Bancos (FEBRABAN) realiza pesquisa anual intitulada "Pesquisa FEBRABAN de Tecnologia Bancária" que ilustra bem o tamanho do setor no Brasil e, principalmente, seu ritmo de crescimento. Publica também um Relatório Anual onde, além de informações económicas, apresenta o resultado das diretrizes da *Global Reporting Initiative (GRI)*[6].

O Banco de Portugal mantém em seu site aplicativo denominado Banco de Portugal - Estatísticas Online (BPstat)[7], que, de maneira eficiente, possibilita a difusão de estatísticas relevantes sobre a economia portuguesa e sobre o setor bancário no país.

O Banco Mundial disponibiliza ferramenta com inúmeras bases de dados, dentre elas bases com informações sobre a indústria financeira mundial[8]. A base de dados intitulada *Global Financial Inclusion* contém 776 indicadores sobre 183 economias, incluindo indicadores de medida de inclusão financeira e sobre como as pessoas guardam, emprestam, fazem pagamentos e a gestão do risco.

Importante ressaltar que o conceito de inclusão financeira é abrangente, conforme está registado no documento Políticas de Inclusão e Formação Financeira [13], resultante do Encontro dos Bancos Centrais dos Países de Língua Portuguesa:

> "O conceito de inclusão financeira envolve um critério não só quantitativo de acesso a produtos bancários, mas também qualitativo sobre a sua adequada utilização. A inclusão financeira é, em primeiro lugar, entendida como o acesso a uma conta bancária (bancarização), porque a posse de uma conta é um requisito essencial para aceder a outros produtos e serviços financeiros. Mas o conceito de inclusão financeira deve ainda abarcar o acesso a outros produtos e serviços financeiros e o seu uso efetivo por parte do consumidor. Ter uma conta bancária não significa, por si só, uma utilização adequada da conta e dos produtos e serviços financeiros associados." [13, p. 12]

Assim sendo, para efeito deste estudo, a informação mais relevante é a quantidade de contas ativas, ou seja, bancarização, uma vez que, objetivamente, este número permite aferir o percentual da população que está gerando transações nas instituições financeiras.

[6] https://www.globalreporting.org
[7] https://bpstat.bportugal.pt/?mlid=1327
[8] http://databank.worldbank.org/data/home.aspx

Utilizando relatório e dados destas fontes de informações, é possível traçar um perfil do setor bancário, incluindo o volume de transações processadas, por canal de atendimento.

Em 2020, 178 milhões de consumidores brasileiros mantiveram relacionamento com as instituições financeiras [55], representando algo em torno de 80% da população, contudo relatórios internacionais indicam que até 2017 somente 70% da população brasileira possuía conta corrente num banco [120]. Um percentual alto para os padrões latino-americanos e mesmo entre os países em desenvolvimento, Brasil, Rússia, Índia, China e África do Sul (BRICS), porém, muito baixo se comparado com países desenvolvidos. Ou seja, existe um potencial de crescimento deste número de clientes, que pode ocorrer rapidamente, considerando que entre 2011 e 2017 este crescimento foi de 15%, porém, entre 2014 e 2017 o crescimento foi de apenas 2%, conforme a Figura 2.8 [120].

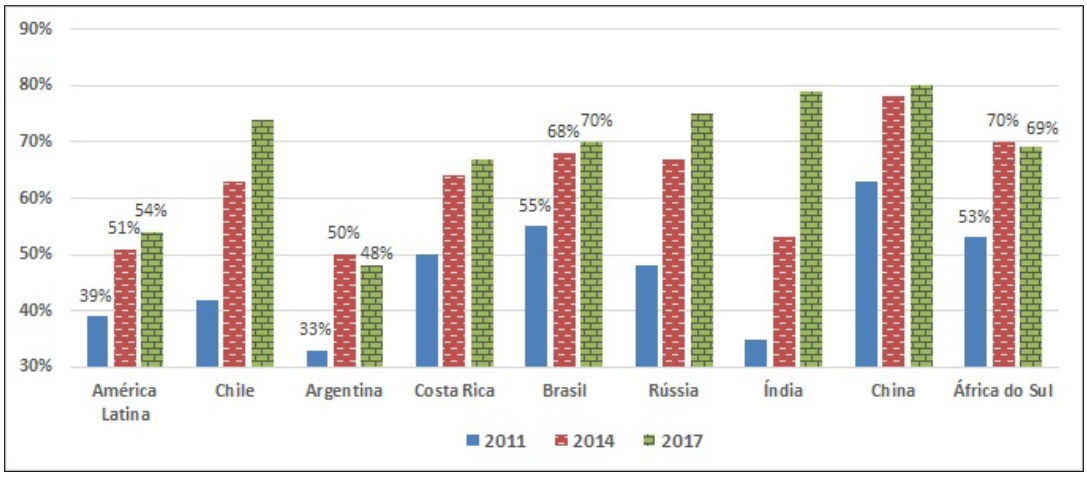


Figura 2.8: Bancarização: Brasil x América latina x BRICS

Este crescimento deve ser considerado significativo pois em 2020 foram realizadas 103.500 milhões de transações nos bancos brasileiros, sendo que esta quantidade cresceu 20% no último ano. Importante salientar que deste montante, 66.400 milhões de transações são passiveis de análise no escopo da tarefa de Anti-Branqueamento de Capitais (ABC) [55].

Em termos de tendência as tradicionais transações com cheques, mais lentas e menos complexas de serem analisadas sob a ótica do ABC, caíram de 1.120 milhão em 2010 para 219 milhões em 2021 (-80%). Em 2020 os canais Internet e *Mobile Banking* confirmaram a preferência no relacionamento com os bancos, com 66% das transações sendo realizadas por estes canais, em detrimento do relacionamento presencial nas agências (9%), onde procedimentos de ABC são mais fáceis de serem aplicados [55].

Segundo dados do Banco Mundial, Portugal apresenta um ótimo índice de bancarização, 92% em 2017, porém abaixo da média da Área do Euro e de outros países desenvolvidos como França e Alemanha. Isso significa que, também em Portugal, existe um potencial de crescimento deste número de clientes, considerando que entre 2011 e 2017

este crescimento foi da ordem de 11%, enquanto a Área do Euro cresceu 5%, conforme a Figura 2.9 [120].

O Banco Mundial disponibiliza a base de dados, sem tecer comentários ou análises de todos os valores encontrados. Relatórios especiais são produzidos, porém, o tema bancarização não foi abordado por nenhum desses relatórios, por isso, não existe uma explicação oficial para a queda no percentual da população com conta corrente em países como Espanha, França e Reino Unido.

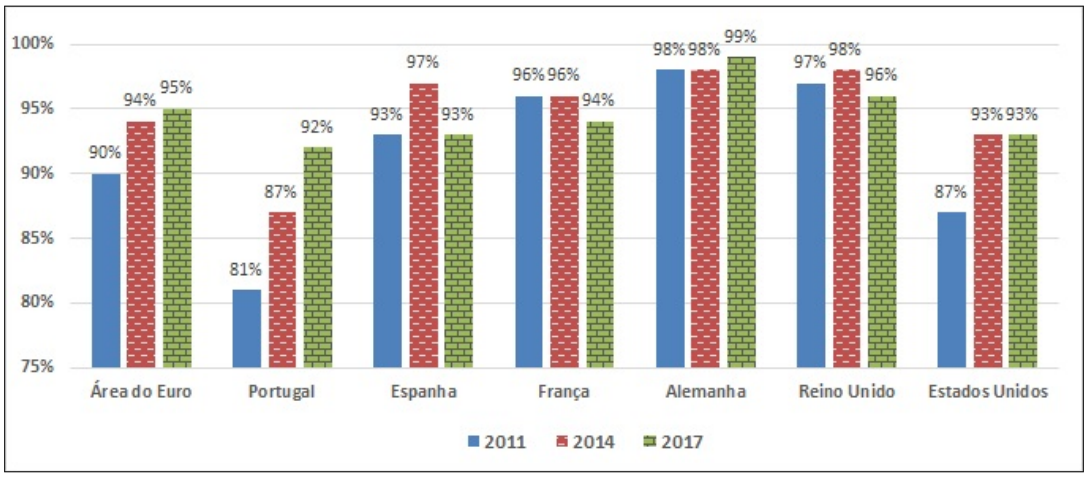


Figura 2.9: Bancarização: Portugal x Área do Euro x USA

Para os padrões europeus um crescimento de 11% é muito significativo já que em 2018 foram realizadas em torno de 2.277 milhões de transações, com impacto financeiro, nos bancos portugueses. Todas essas transações são passiveis de análise no escopo da tarefa de ABC [14].

Relativamente à utilização de cheques, em Portugal a redução foi semelhante à do Brasil e ficou em torno de 84% entre 2010 e 2021. O canal Multibanco em 2021 representou em torno de 90% do total das transações realizadas, relembrando que neste tipo de canal os procedimentos de ABC são mais difíceis de serem aplicados [14].

Tendo por base os percentuais de crescimento citados, e revisitando a Figura 2.7, é possível perceber que o crescente aumento no volume de transações geradas pelos sistemas de negócio, que provocam, consequentemente, um maior volume de transações com indícios de irregularidades, resultando na necessidade de mais analistas responsáveis pela atividade de verificação dessas suspeitas. Um número insuficiente de analistas provoca atrasos na comunicação aos órgãos reguladores tornando o processo ineficiente quanto a um rápido combate e completamente ineficaz quanto a tarefa de prevenção.

2.2.3 Um Possível Risco Sistémico

Fica evidente que o sucesso do atual processo depende, basicamente, de dois fatores: uma análise bem realizada por parte dos analistas humanos e parâmetros bem definidos

visando uma melhor seleção dos casos a serem analisados.

Ampliando a visão de como o processo ocorre em uma instituição financeiro (Figura 2.7) e considerando o conjunto das instituições financeiras que integram o sistema financeiro, bem como os órgãos reguladores, a Figura 2.10 busca mostrar como ocorre a retroalimentação no processo. Todas as instituições financeiras reportam as transações suspeitas às autoridades competentes (COAF, UIF ou PGR), quer estas suspeições tenham sido descobertas pelos processos internos quer tenham tido origem em uma denúncia (chamada na figura de sinalização externa). Estas sinalizações externas também podem ser encaminhadas para o órgão regulador (BACEN ou BdP).

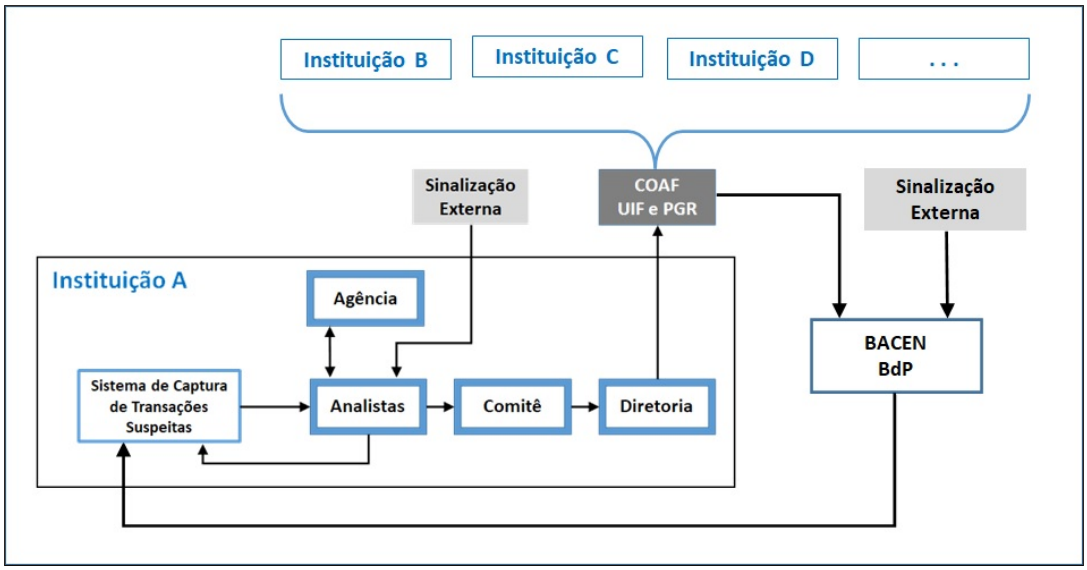


Figura 2.10: Retroalimentação do Processo de ABC

Analisando este processo sob a perspetiva da retroalimentação é possível perceber as seguintes situações:

- o aprimoramento dos parâmetros utilizados para verificação e seleção dos casos suspeitos é feito tendo por base as normas estabelecidas e o conhecimento adquirido nas verificações realizadas pelos analistas;
- os analistas só aprofundam a verificação nos casos selecionados pelas rotinas que utilizam os citados parâmetros ou por sinalizações externas (denuncias, ações policiais, trabalhos jornalísticos etc.);
- os Bancos Centrais são responsáveis pela evolução dos normativos, baseando-se nas experiências repassadas pelos organismos de controle e nos casos efetivamente comprovados.

Existe nesta descrição um fluxo pouco favorável a um adequado aprimoramento da etapa inicial, que é a seleção dos casos com indicativos de suspeição da prática criminosa, ponto crucial de retroalimentação do processo. Relatos de participantes do processo dão

conta de que os analistas dedicam cada vez menos tempo à tarefa de extrair dos casos analisados as situações passíveis de utilização no aprimoramento dos parâmetros dos sistemas de captura, em função do crescimento no número de transações sinalizadas para análise. Parâmetros inadequados resultam em sinalizações ineficientes e até omissas, análises demoradas e sem a devida profundidade levam a comunicações de casos suspeitos aquém da realidade de mercado, permitindo a inferência de que os casos relatados não permitem aos órgãos de controle aprimorarem, adequadamente, as normas e recomendações. Criando-se, dessa forma, um círculo vicioso.

Uma situação rotineira que contribui para fragilizar este processo é a dissociação entre os casos comunicados pelas instituições financeiras e o resultado de um possível processo instaurado, ou seja, a comprovação do Branqueamento de Capitais (BC). Processos dessa natureza podem levar anos até sua conclusão e não existe um fluxo de retorno de informação que permita estabelecer uma relação entre a transação comunicada como suspeita e a efetiva comprovação da fraude.

Outro fator que reforça a necessidade de cada instituição financeira adotar um robusto processo de verificação é a capacidade de adaptação do *modus operandi* dos fraudadores. Um bom exemplo dessa capacidade é a manipulação dos valores limites nas transações realizadas. Os bancos centrais de cada país definem um valor na moeda local a partir do qual todas as transações de transferência e/ou saque deve ser comunicado ao órgão fiscalizador. Para fugir a este controle os fraudadores realizam transações com valores inferiores a este limite e utilizam diversas instituições financeiras.

Esta é uma situação que extrapola o âmbito da instituição financeira e compete ao órgão supervisor proceder esta verificação. No entanto, para viabilizar este procedimento seria necessário que as instituições financeiras informassem todas as transações realizadas e não somente aquelas superiores a um limite. Fica claro, contudo, que este é um procedimento inviável em função do volume de informação. Resta, então, como alternativa que cada instituição financeira aprimore seus processos, permitindo um conhecimento e acompanhamento do perfil de cada cliente, visando mitigar este risco.

Apesar de importantes, considerando o reflexo das particularidades de cada país, o aprimoramento baseado em casos locais comprovados é apenas uma das fontes de informações que permitem aos organismos de controle produzirem normas e recomendações. O atual nível de cooperação internacional, coordenado pelas já citadas instituições como FATF, GAFI e Grupo de Egmont, permitem uma frequente atualização da base legal e das recomendações que resultam em parâmetros para identificação de transações suspeitas.

Buscando mitigar esse erro sistémico, as instituições financeiras buscam constantemente novas formas de melhor identificar as transações suspeitas, no entanto, a prática de mercado e os estudos publicados mostram um uso constante de um fluxo baseado, quase exclusivamente, na política Know Your Customer (KYC) para combater o BC [57, 40]. Torna-se, portanto, necessário que novos mecanismos de análise que permitam uma rápida, de preferência imediata, sinalização de transações suspeitas, com base nas mais recentes normas e recomendações publicadas ([1, 2, 3, 4, 5, 6]).

2.3 FAIS - Primeira Ferramenta de Apoio ao Anti-Branqueamento de Capitais

Uma das ferramentas mais conhecidas na área do Anti-Branqueamento de Capitais (ABC) é o *FinCEN Artificial Intelligence System (FAIS)* [113], desenvolvido e utilizado pelo *Financial Crimes Enforcement Network (FinCEN)*[9], órgão do Departamento do Tesouro dos Estados Unidos. A missão do FinCEN é estabelecer, implementar e supervisionar medidas de prevenção e combate ao crime de Branqueamento de Capitais (BC) e financiamento do terrorismo e utiliza o FAIS desde 1993 para avaliar relatórios e identificar potenciais casos.

O FAIS analisa e estabelece ligações entre os relatórios fornecidos pelas instituições financeiras americanas. Tais relatórios contêm informações sobre transações suspeitas, bem como transações realizadas acima do valor limite estabelecido em lei. É um sistema baseado em regras, utiliza técnicas de *data mining* e *link analysis* e seu ponto chave é o "*suspicion score*", que pontua os vários tipos de transação e atividade ilícita, porém, oferece grande ênfase na visualização do resultado encontrado [21].

> *O FAIS foi baseado em análise de formulários e não apresenta indicação conclusiva sobre as análises realizadas, todos os resultados estatísticos e de ponderação de risco são submetidos aos analistas humanos para decisão.*
>
> *(Nota A)*[10]

Em 1995, a Subcomissão Permanente de Investigações do Comité do Senado Americano para Assuntos Governamentais solicitou ao *Office of Technology Assessement (OTA)*, escritório do Congresso Americano que funcionou de 1972 a 1995[11], para avaliar proposta de utilização de técnicas de pesquisa baseadas em inteligência artificial visando monitorizar, de forma totalmente automatizada, o tráfego e transferências bancárias com o propósito de reconhecer transações suspeitas.

Em 1998, relatório do *United States General Accounting Office (GAO)*[12] registou que o FinCEN decidiu parar alguns serviços preventivos de combate ao BC e, principalmente, também decidiu não implementar novos produtos utilizando técnicas de Inteligência Artificial (IA). Os principais motivos identificados para esta decisão foram: a redução de pessoal apto a desenvolver estas atividades; e porque as agências de informações, muitas vezes, não tomavam qualquer ação diante das sinalizações efetuadas [102]. Auditoria realizada em 2001 pelo *Office of Inspector General (OIG)*[13] atestou que somente um *Intelligence Research Specialist*[14] estava alocado para trabalhar com o FAIS, como resultado

[9] https://www.fincen.gov/

[10] As Notas apresentadas neste trabalho são comentários do autor sobre o item abordado. O principal objetivo é sinalizar os aspetos positivos que foram considerados no desenvolvimento do trabalho.

[11] http://ota.fas.org/

[12] http://www.gao.gov/

[13] https://www.oig.doc.gov

[14] https://www.fincen.gov/careers/intelligence_research.html

desta auditoria foram alocados mais seis especialistas, contudo, não há informações sobre novas versões do sistema [92].

A avaliação do OTA concluiu que "o conceito original na sua formulação mais simples – o monitoramento do tráfego de transferência bancária, de forma contínua e em tempo real, usando técnicas de inteligência artificial – não é viável". Contudo, ponderou que existiam formas em que a tecnologia da informação podia ser utilizada para apoiar e reforçar a aplicação da lei contra o BC [94]. O capítulo quatro intitulado *Technologies for Detecting Money Laundering*, que integra o documento [94] sugere algumas técnicas tais como: *knowledge acquisition*, *machine learning*, *clustering*, *knowledge sharing* e *data transformation*.

> *Apesar do parecer final do OTA, à época, ser contrário à utilização de IA, as técnicas alternativas sugeridas eram viáveis e a decisão de não evoluir o sistema não parece ter sido técnica, mas sim gerencial.*
>
> *(Nota B)*

Capítulo 3

Técnicas e Abordagens para apoio ao Anti-Branqueamento de Capitais

O tema Anti-Branqueamento de Capitais (ABC) não é dos mais preferidos para aplicação de estudos académicos e soluções inovadoras. Um dos motivos talvez seja a dificuldade na obtenção de dados reais das instituições financeiras e, principalmente, a necessidade de manutenção de sigilo sobre os dados obtidos.

O objetivo deste capítulo é apresentar um levantamento das abordagens que têm sido utilizadas no desenvolvimento de sistemas de apoio ao combate ao Branqueamento de Capitais (BC), bem como uma visão geral das técnicas que lhes são subjacentes, tais como: mineração de dados, aprendizado de máquina, agentes inteligentes e sistemas multi-agentes. No final de cada secção, encerrando os principais temas são referenciados os trabalhos identificados e que têm relação com os temas descritos.

3.1 Mineração de Dados e Aprendizado de Máquina

Barateamento dos meios de armazenamento, popularização de equipamentos dotados de sensores de captura, incremento constante na velocidade de transmissão de dados, são fatores que determinaram o grande aumento no volume de dados armazenados. Apesar de ter ocorrido esta mudança de cenário nas técnicas e meios de coleta e armazenamento de dados, indo do armazenamento caro e uso pontual até a integração e ampla possibilidade de utilização nos mais diversos setores, informações úteis para as pessoas, negócios e pesquisas carecem de ser descobertas. Técnicas de mineração de dados e aquisição de conhecimento são as mais utilizadas para esta descoberta de informações.

3.1.1 Introdução

O processo decisório nos mais diversos setores necessitam de respostas para perguntas cada vez mais complexas e que, na maioria das vezes, não podem ser respondidas com meras consultas utilizando linguagens de banco de dados. Perguntas complexas e funda-

mentais para a prosperidade ou sobrevivência de um negócio, tais como: Qual produto de alta lucratividade venderia mais com a promoção de um item de baixa lucratividade, analisando os dados dos últimos dez anos de vendas? ou quais são os clientes potenciais para praticar um ilícito financeiro? Desta forma, torna-se necessária a utilização de ferramentas de análise e extração de conhecimento para responder estas perguntas [104].

A estratégia básica utilizada é a criação de bases de dados que contenham dados limpos, agregados e consolidados, que possam ser analisados por ferramentas capazes de realizar consultas complexas nestas bases multidimensionais. Analisar e compreender grande volume de dados é uma deficiência reconhecida, por isso, estudos visando ao desenvolvimento de tecnologias de extração automática de conhecimento de Bases de Dados continuam sendo desenvolvidos nesta área denominada *Knowledge Discovery in Databases (KDD)* ou Mineração de Dados (MD).

3.1.2 Etapas do processo de mineração de dados

Os autores não divergem sobre o processo de MD ser dividido em fases, contudo, eles divergem quanto ao número de fases. Fayyad et al. em [54] propôs um processo com cinco fases, Weiss e Indurkhya [124] apresentou fluxo com quatro fases, Rezende et al. [104] utiliza três fases, numa publicação mais recente, Tan et al. [118] reforça o modelo em três fases. A Figura 3.1 apresenta o processo com três fases principais, baseado em [118], mostrando os principais passos a serem executados.

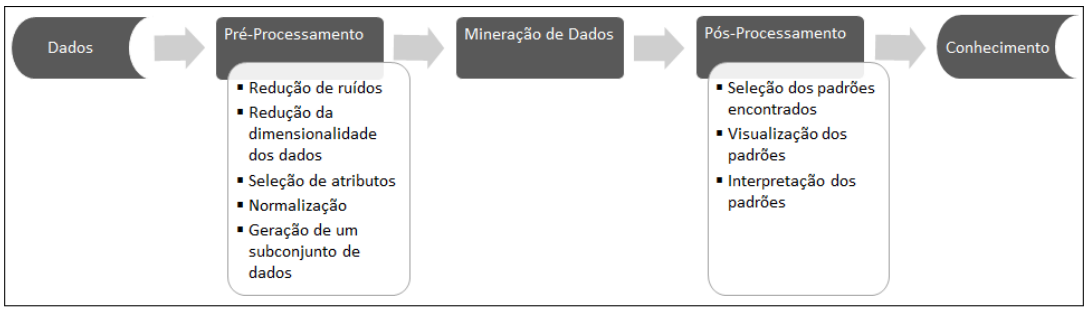


Figura 3.1: Processo de Mineração de Dados

A forma de armazenamento e a localização dos dados podem ser as mais variadas, por isso, o **pré-processamento** é chamado de fase da transformação dos dados. Envolve o agrupamento e escolha da massa de dados a ser minerada, onde a extração será realizada, busca assegurar a qualidade dos dados envolvidos na mineração, realizando operações básicas como a remoção de ruídos e exigindo decisões estratégicas em casos de omissão de dados, além, quando existente, da redução das dimensões destes dados. Nesta fase também ocorre a procura por atributos úteis nos dados tendo em consideração os objetivos a que se destinam e buscando à redução do número efetivo de variáveis. Um subconjunto de dados é então gerado. Na fase seguinte a **mineração de dados** é efetivamente iniciada, com a escolha do método e do algoritmo mais compatível com o objetivo da extração, a

fim de encontrar padrões nos dados que sirva de subsídios para descobrir conhecimentos ocultos. A fase **pós-processamento** é a fase de consolidação do conhecimento descoberto, incorporação deste conhecimento no sistema ou elaboração de relatórios para as partes interessadas, verificação e resolução de potenciais conflitos com conhecimento tido como verdadeiro [104, 118]. Estas fases foram aplicadas na execução deste trabalho e estão descritas no Capítulo 4 - pág. 45.

Em Camilo [27] foi apresentada uma classificação de MD, conforme sua capacidade de realizar as tarefas, seguindo proposta inicial de [77], publicada na primeira edição do livro:

- Descrição (*Description*) é a tarefa utilizada para descrever os padrões e tendências revelados pelos dados. Geralmente estas descrições sugerem explicações para os padrões e tendências descobertos;

- Estimação (*Estimation*) ou Regressão (*Regression*) é usada quando o registo é identificado por um valor numérico e não um categórico. Assim, pode-se estimar o valor de uma determinada variável analisando-se os valores das demais. A tarefa de estimação pode ser usada, por exemplo, para estimar a pressão ideal de um paciente baseando-se na idade, sexo e massa corporal;

- Classificação (*Classification*) é similar a estimação, porém, é utilizada quando a variável alvo é categórica. É uma das tarefas mais comuns, visa identificar qual classe um determinado registo pertence. Nesta tarefa, o modelo analisa o conjunto de registos fornecidos, com cada registo já contendo a indicação à qual classe pertence, a fim de 'aprender' como classificar um novo registo. A tarefa de classificação pode ser usada por exemplo para: determinar quando uma transação de cartão de crédito pode ser uma fraude; identificar a que grupo de risco um determinado cliente pertence;

- Predição (*Prediction*) esta tarefa é similar às tarefas de classificação e estimação, porém ela visa descobrir o valor futuro de um determinado atributo. Exemplos: predizer o valor de uma ação três meses adiante; predizer o percentual que será aumentado de tráfego na rede se a velocidade aumentar;

- Agrupamento (*Clustering*) a tarefa de agrupamento busca identificar e aproximar os registos similares. Um agrupamento (ou *cluster*) é uma coleção de registos similares entre si, porém diferentes dos outros registos nos demais agrupamentos. Esta tarefa difere da classificação pois não necessita que os registos sejam previamente categorizados. O agrupamento não tem a pretensão de classificar, estimar ou predizer o valor de uma variável, ela apenas identifica os grupos de dados similares. Exemplos: segmentação de mercado para um nicho de produtos; separando comportamentos transacionais que podem ser suspeitos;

- Associação (*Association*) consiste em identificar quais atributos estão relacionados. Apresentam a forma: SE atributo X ENTÃO atributo Y. É uma das tarefas mais conhecidas devido aos bons resultados obtidos. Alguns exemplos: determinar os casos onde um novo medicamento pode apresentar efeitos colaterais; identificar os usuários de planos que respondem bem a oferta de novos serviços.

3.1.3 Aprendizado de Máquina

Definições informais, em alguns documentos ou nas descrições de ferramentas de mercado, colocam Mineração de Dados (MD) como uma aplicação de Aprendizado de Máquina (AM). É possível que esta visão exista em função de que boa parte das ferramentas utilizadas em MD estão baseadas em algoritmos de AM. Por outro lado, MD atualmente aparece como forte indutora para AM, assim sendo, "é inegável a existência de uma forte sinergética sobreposição entre as duas áreas" [100, p. 2].

Para Monard and Baranauskas [90] "um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas por meio da solução bem-sucedida de problemas anteriores". Estes sistemas podem ser classificados quanto à linguagem de descrição, modo, paradigma e forma de aprendizado utilizado. Importante observar que não existe um método de aprendizado dito de "propósito-geral", ou seja, não existe uma solução que apresente o melhor desempenho para todos os problemas. Cada método tem a sua utilidade comprometida pelas suas suposições e particularidades, e cada aplicação requer compreensão das limitações dos diversos algoritmos de AM e utilizar abordagem que permita avaliar conceitos por eles induzidos [90, 100].

Utilizar o conhecimento já adquirido para obter novo conhecimento é um processo natural do aprendizado. Quanto mais inferência for aplicada no processo de aquisição deste novo conhecimento, menos a fonte de conhecimento ou o ambiente externo precisará atuar. No aprendizado humano cinco estratégias podem ser enumeradas, segundo o grau de complexidade de inferência: hábito, instrução, dedução, analogia e indução. A primeira estratégia apresenta menor complexidade de inferência, ao passo que a estratégia indutiva exige maior esforço para o aprendizado [88, 89].

Em Metz [88], citando Michalski et al. [89], é esclarecido que no aprendizado por hábito, todo conhecimento é transmitido do instrutor para o aprendiz, o qual não realiza nenhuma inferência sobre as informações fornecidas, apenas as memoriza. No aprendizado por instrução, o aprendiz adquire conceitos de uma fonte (professor e livros textos, por exemplo) mas não transfere diretamente a informação recebida para a memória. Essa estratégia engloba a seleção dos fatos mais importantes e a transformação da informação fonte em formas mais apropriadas. No aprendizado por dedução, o aprendiz adquire um conceito por meio da dedução sobre o conceito já conhecido, ou seja, o conhecimento aprendido é o resultado de uma transformação sobre um conhecimento que o indivíduo possui a priori. O aprendizado por analogia é caracterizado quando o aprendiz modifica conceitos previamente adquiridos para aprender novos conceitos. Assim, ele não cria

regras, mas adapta regras existentes para que possam descrever o novo conceito.

No aprendizado indutivo é utilizada uma forma de inferência lógica que permite obter conclusões genéricas sobre um conjunto particular de exemplos. O raciocínio é originado de um conceito específico para um conhecimento genérico, ou seja, da parte para o todo. Na indução, um conceito é aprendido efetuando a inferência indutiva sobre os exemplos apresentados – construção de uma hipótese. Essa inferência pode ser verdadeira ou não [90]. Basicamente o que diferencia os tipos de aprendizado indutivo é a existência ou não de um atributo especial chamado classe que rotula os registos do conjunto de dados utilizado. Quando esta classe existe o aprendizado é dito supervisionado; quando apenas parte dos exemplos apresenta este atributo o aprendizado é chamado semissupervisionado; e quando esses rótulos não existem o aprendizado é não-supervisionado [88].

Outro tipo de aprendizado indutivo é chamado por reforço, que consiste no treinamento de modelos de aprendizado de máquina para tomada de decisões. O aprendizado consiste em atingir um objetivo em um ambiente incerto e complexo. No aprendizado por reforço, o sistema enfrenta uma situação e busca encontrar uma solução para o problema aprendendo com tentativas e erros. O processo de aprendizado recebe recompensas ou penalidades pelas ações que executa. Portanto, assim como no aprendizado humano, o aprendizado indutivo por reforço selecionam ações que lhes proporcionariam a maior recompensa [72]. A Figura 3.2 mostra a hierarquia do aprendizado indutivo.

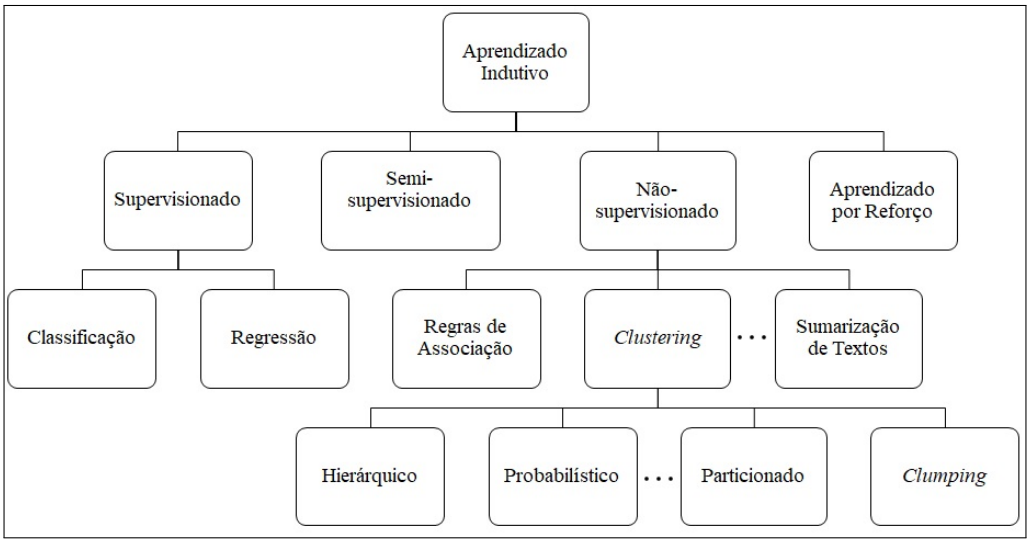


Figura 3.2: Hierarquia do Aprendizado Indutivo

3.1.4 Aplicações para o Anti-Branqueamento de Capitais

A mineração de dados tem sido uma das principais técnicas utilizadas na tentativa de identificar padrões ou anomalias que possam indicar a prática do crime de branqueamento de capitais. Assim sendo, existem vários trabalhos publicados com propostas nesta área.

Zhang et al. [132] discretizou um conjunto de dados para construção de *clusters*. Os recursos são mapeados para N+2 espaço dimensional Euclidiano, sendo N dimensões do cliente, uma dimensão de tempo e uma dimensão para transações. A frequência das transações de cada cliente é projetada numa linha de tempo discretizada, formando um histograma. Os *clusters* são então criados (utilizando o algoritmo K-means) com base nos segmentos no histograma. Análises de correlação locais e globais são então aplicadas para detetar padrões suspeitos.

> *Uma boa abordagem para a análise de comportamentos individuais e/ou comportamentos de grupo, examinando as suas operações para detetar comportamentos suspeitos relacionados com picos anormais no histograma. No entanto, quando é necessário analisar grande quantidade de clientes e transações durante um longo período de tempo, pode ficar difícil detetar casos suspeitos, pois podem existir poucos ou nenhum pico no histograma.*
>
> *(Nota C)*

Kingdon [74] propôs a utilização de técnicas de Inteligência Artificial (IA) e a utilização de máquina de vetor de suporte (*Support Vector Machine (SVM)*) numa perspectiva diferente daquela até então utilizada, ou seja, montar o perfil histórico dos clientes e buscar utilizações fora do padrão ao invés de focar em comportamento suspeito baseado apenas em perfil cadastral.

> *Independente da tecnologia utilizada, esta proposta fortalece o conceito da política* Know Your Customer (KYC) *e torna-se ainda mais útil se aplicada de forma incremental e constante.*
>
> *(Nota D)*

A implementação realizada por Kingdon [74] é uma extensão de uma *SVMs* apresentada por Scholkopf et al. [111], na qual é proposta a utilização de uma matriz com grande dimensionalidade. Teoricamente um fator positivo desta abordagem é ela poder lidar com conjuntos de dados heterogéneos, contudo, o tamanho da matriz deixou muitas dúvidas quanto ao desempenho [79, 49].

> *O questionamento apresentado sobre esta proposta continua válida, considerando que não foi possível localizar testes mais exaustivos.*
>
> *(Nota E)*

Tang and Yin [119] propuseram outra extensão da *SVM* para analisar as transações dos clientes e detetar comportamento fora do padrão. É apresentado uma combinação de um *kernel* com melhoramentos do *Radial Basis Function (RBF)* [111] com definição de distâncias distintas [125] e algoritmos SVM supervisionados e não-supervisionados.

> *Uma vantagem desta abordagem é o fato dela conseguir lidar com conjuntos de dados heterogéneos, porém, a avaliação de desempenho foi feita apenas com dados de simulação.*
>
> *(Nota F)*

Combinar a técnica de *clustering* com *Multilayer Perceptron (MLP)* foi a proposta de Le-Khac et al. [80]. Um processo de *clustering*, aplicando o algoritmo K-means, é utilizada para detetar casos suspeitos de Branqueamento de Capitais (BC). Esta técnica baseia-se em duas características principais (fundo de investimento e investidor) que, em seguida, são usados como entrada do processo de formação de um Multilayer Perceptron (MLP).

> *Os resultados apresentados mostram que a sua abordagem é eficiente. No entanto, o número de características e o número de padrões de treinamento utilizados foi muito pequeno e isso pode afetar a precisão.*
>
> *(Nota G)*

Outros métodos estatísticos também foram propostos como por Liu and Zhang [84], que utiliza *scan statistics*, onde as transações realizadas num período de tempo são selecionadas aleatoriamente e agrupamentos incomuns são buscados.

> *Adequado para trabalhos de auditoria em função da aleatoriedade imposta ao procedimento, no entanto, para um processo quotidiano nenhuma transação pode ser desprezada.*
>
> *(Nota H)*

Le-Khac and Kechadi [79] apresentam um estudo de caso em que aplicam uma solução para geração de uma base de conhecimento, combinando técnicas de mineração de dados, *clustering* utilizando K-means, redes neurais e algoritmos genéticos para detetar padrões de BC.

> *Analisando somente por este documento, a solução pode apresentar um baixo custo-benefício, considerando a alta complexidade da implementação proposta.*
>
> *(Nota I)*

3.2 Sistemas Baseados em Agentes aplicados ao ABC

3.2.1 Introdução

A área dos agentes apresenta uma grande interdisciplinaridade, como mostra Helder Coelho em [36, p. 46] posicionando os agentes da seguinte forma:

> "a área dos agentes está colocada nas Ciências da Computação, mais propriamente no campo do Processamento Inteligente da Informação, a par da Descoberta do Conhecimento, da Gestão do Conhecimento, da Aprendizagem Mecânica e do Raciocínio Autonómico, ou seja, numa região onde a Inteligência Artificial se cruza com os Sistemas Distribuídos, o Reconhecimento de Padrões ou a Robótica."

Quer seja pela evolução da informática ou por motivações científicas, desde sempre o domínio da Inteligência Artificial Distribuída (IAD) foi dividido em duas abordagens distintas: Resolução Distribuída de Problemas (RDP) e Sistemas Multi-Agente (SMA). É possível considerar a abordagem RDP como sendo "um cruzamento das técnicas de IA e de sistemas distribuídos, que aplicam as técnicas de coordenação e sincronização desenvolvidas no segundo para integrar sistemas desenvolvidos conforme modelos do primeiro"[66, p. 280]. Um SMA é um sistema composto por vários agentes inteligentes que interagem para resolver problemas que vão além das capacidades ou conhecimentos individuais de cada agente [47].

Nos últimos anos, observou-se uma mudança de estratégia na IA enquanto ciência, ela procurou se utilizar das teorias existentes como base, em vez de procurar por soluções completamente novas. Elementos como Internet, mecanismos de pesquisa e grandes volumes de dados permitiram este novo caminho. Atualmente agente inteligente é o conceito mais comumente aceito, "no qual as abordagens simbólicas e conexionistas podem trabalhar de forma colaborativa para a resolução de problemas através de um sistema computacional"[60, p. 17].

> *O sistema idealizado neste trabalho guarda afinidade com este conceito quando utiliza agentes inteligentes aliados a técnicas de aquisição de conhecimento para lidar com grande base de dados na busca de solução para alguns problemas enfrentados pelo Anti-Branqueamento de Capitais (ABC).*
>
> *(Nota J)*

3.2.2 Agentes

Sobre o surgimento dos agentes Jaime Sichman e Helder Coelho [115, p. 5] relatam que:

> "O campo de agentes nasceu nos EUA no início dos anos 80, com o Workshop de Inteligência Artificial Distribuída (DAI), na Europa, no final dos anos 80, com os Agentes Autónomos de Modelagem em um workshop com o Multi-Agent World (MAAMAW) e na Orla do Pacífico, no início dos anos 90, com o Workshop Japonês sobre Multi-Agente e Computação Cooperativa (MACC). Posteriormente, enquanto as soluções multi-agente começam a surgir, esses três eventos foram federados na Conferência Internacional sobre Sistemas Multi-Agente (ICMAS), cuja primeira ocorrência foi em San Francisco, em 1995. Nos anos 90, surgiram outras duas conferências e workshops: a Conferência Internacional sobre Agentes Autónomos (AA) e o Workshop Internacional sobre Teorias, Arquiteturas e Idiomas de Agentes (ATAL). A partir de 2002, estes últimos foram fundidos com o ICMAS, levando à atual Conferência de Agentes Autónomos e Sistemas Multi-Agente (AAMAS). O supervisor desta conferência, o conselho internacional da IFAAMAS, é agora o órgão líder de todo o campo e da comunidade de pesquisa de agentes e sistemas multi-agentes."

Uma das definições mais conhecidas de agente coloca-o como um sistema de computador localizado em algum ambiente e que é capaz de ações autónomas sobre este ambiente visando atingir um objetivo recebido [129]. Em princípio, o termo *computer system* parece muito abrangente para definir um agente, porém, ela pode ser esclarecida com uma definição mais detalhada de SMA, que será feita na secção 3.2.3. No entanto, vários autores defendem que agentes são sim um tipo especial de sistema computacional, neste sentido um SMA seria um conjunto de sistemas.

Mais direta é a definição de Costa and Simões [38, p. 28] que apresenta agente como sendo toda a entidade capaz de interagir com o ambiente guiado, em geral, por objetivo. Complementam definindo que a estrutura de um agente é algo simples:

> "tem um mecanismo que lhe permite recolher informações do ambiente (perceção), mecanismos que lhe permitem atuar sobre o ambiente (ação) e processos que lhe permitem definir qual a melhor ação a realizar (decisão). Os processos de decisão serão tanto mais sofisticados quanto mais complexa for a tarefa e/ou o ambiente."

Em termos estruturais um agente é realmente algo simples, a Figura 3.3 mostra uma arquitetura de mais alto nível para um agente.

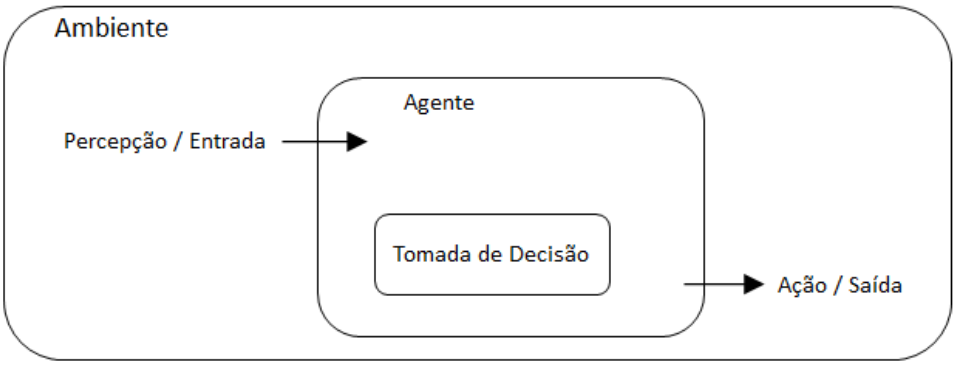


Figura 3.3: Arquitetura de um Agente

3.2.2.1 Propriedades

As definições existentes para agentes são genéricas e podem representar um elemento físico ou um elemento virtual, contudo, todos os estudos afirmam que existe um conjunto de propriedades básicas e atributos que caracterizam um agente. São várias as relações dessas propriedades e atributos, apresentando pequenas variações de definição ou nomenclatura, sendo as seguintes as principais [66, 130, 129]:

- Autonomia de Decisão - capacidade de analisar uma situação, gerar alternativas de atuação e escolher a situação que melhor atende seus objetivos. Em certos casos, o agente não reconhece o cenário de atuação, mas tem capacidade de escolher uma experiência prévia semelhante e adaptar a solução ao novo cenário;

- Autonomia de Execução - capacidade de operar no ambiente sem intervenção de outro agente (geralmente humanos);
- Competência para Decidir - capacidade de configurar sua atuação sem intervenção externa;
- Existência de uma Agenda Própria - capacidade de criar uma lista de objetivos que concretizem suas metas;
- Reatividade - capacidade de reagir às mudanças do ambiente a partir do reconhecimento de um contexto conhecido;
- Adaptabilidade - capacidade do agente de adaptar seu processo de decisão frente a situações desconhecidas;
- Mobilidade - capacidade do agente de mover-se e ser executado em outras plataformas;
- Personalidade - capacidade do agente de personificar-se, utilizando recursos que lembrem características humanas como a emoção ou o humor;
- Interatividade com o Usuário - capacidade de interagir com usuários e, considerando os possíveis mal-entendidos, reagir às falhas de comunicação de maneira aceitável;
- Ambiente de Atuação - caracteriza o local onde o agente vai atuar, isto é, em ambientes fechados (*desktop*) ou abertos (Internet); e
- Comunicabilidade - capacidade de interagir com outros agentes computacionais para a obtenção de suas metas. Quando mais de um agente atua, seja de modo competitivo ou cooperativo, configura-se um ambiente de atuação multi-agentes.

3.2.2.2 Ambientes

Os ambientes de atuação dos agentes podem ser classificados em ambiente físico ou ambiente de software. Como exemplo de atuação nestes ambientes é possível citar os robôs e a Internet, respetivamente. Russell and Norvig [107] apresenta uma classificação detalhada e longa, Costa and Simões [38] uma classificação mais pragmática com apenas três opções que parecem ser suficientes, contudo, na relação a seguir foi acrescentado o último aspeto deixando a relação com quatro eixos de classificação, isto porque não foi possível encontrar esta última característica nos três itens anteriores:

a) Acessíveis ou não: um ambiente será acessível se o agente puder retirar do ambiente toda a informação que necessita para determinar a melhor ação. O caso clássico de ambientes acessíveis é o que envolve jogos abertos como o xadrez;

b) Deterministas ou não: quando a evolução do ambiente não pode ser determinada de forma única a partir da situação corrente e da ação do agente sobre o ambiente, o ambiente diz-se não determinista. Um caso típico é um sistema de diagnóstico médico, em relação ao qual o comportamento do ambiente (que inclui o paciente e o tratamento) não evolui de forma determinista. Para uma mesma doença, nem todos os pacientes reagem da mesma maneira ao mesmo tratamento;

c) Estáticos ou não: um ambiente diz-se estático se não se altera enquanto o agente está a decidir a ação a executar. Um sistema de previsão do tempo tem de lidar com um ambiente que não é estático, mas sim dinâmico;

d) Agente Único ou não: ambientes com um único agente são aqueles onde somente um agente está colocado para solução do problema. Num ambiente multi-agente mais de um agente está colocado.

Os ambientes mais complexos são aqueles que são inacessíveis, não deterministas, dinâmicos e multi-agentes, enquanto que ambientes acessíveis, deterministas, estáticos e com agente único oferecem as situações mais favoráveis para o agente.

3.2.2.3 Arquitetura

Considerando-se que um agente é um tipo especial de sistema computacional é possível classificá-lo conforme sua representação interna do ambiente. Um agente pode conter um modelo de representação interna do ambiente e dos outros agentes baseado segundo um eixo cognitivo, onde a representação é baseada em estados mentais, possuindo um modelo racional de decisão (agente cognitivo) ou simplesmente agir baseado em reações aos estímulos provocados pelo ambiente (agente reativo) [66, 38].

A arquitetura de um agente, portanto, deve determinar como ele pode ser decomposto em módulos e como estes módulos devem interagir. Este conjunto de módulos e suas interações descrevem como os dados recebidos do ambiente e o estado interno do agente determinam suas ações [130]. Os vários modelos de arquitetura propostos e em uso no mercado adota, também, uma arquitetura denominada híbrida, que apresenta comportamentos inerentes tanto a arquitetura reativa como a cognitiva.

Os agentes reativos são baseados em modelos de organização biológica ou etológica (formigas, cupins, abelhas etc.), seu modelo de funcionamento é formado por um par estímulo-resposta (ação-reação). Nesse modelo não há uma representação simbólica explícita do ambiente, os agentes não têm capacidade de realizar raciocínios complexos e não possuem qualquer tipo de histórico das ações passadas [116, 25].

Os agentes cognitivos utilizam modelos de representação interna do ambiente e dos outros agentes baseado em estados mentais. Permitem a manutenção de histórico das interações e ações passadas, como uma memória, sendo assim capazes de planear ações futuras. As duas formas mais populares de modelagem de um agente cognitivo são a modelagem de agentes cognitivos baseados em lógica e a modelagem de agentes cognitivos

com crenças, desejos e intenções. No modelo baseado em lógica, os estados internos do agente são determinados através de fórmulas lógicas e seu comportamento é determinado pelas regras de dedução e pelo histórico de ações passadas. O segundo modelo, conhecidos como agentes *Belief-Desire-Intention (BDI)*, é baseado na teoria do raciocínio prático humano, desenvolvida, pelo filósofo Michael Bratman [24].

3.2.2.4 Arquitetura BDI

Na arquitetura BDI os estados mentais do agente representam um papel central. O processo de decisão do agente consiste em formar novos desejos (subobjetivos) com base nas crenças, desejos e intenções existentes no momento, como etapa para atingir o objetivo final. Normalmente, esse processo é formado por duas etapas: geração de opções e filtragem. A primeira etapa consiste na escolha de desejos levando em conta as crenças e intenções conhecidas, enquanto a filtragem tem por objetivo escolher a melhor alternativa gerada pela etapa anterior [23, 128]:

- Crenças - as crenças de um agente referem-se ao que o agente acredita ser possível em determinado momento, e descrevem sua perspetiva sobre o estado do ambiente e dos outros agentes;
- Desejos - são todos os possíveis estados existentes para para que o agente alcance seus objetivos, podem ser considerados subobjetivos;
- Intenções - um conjunto de ações ou tarefas que o agente selecionou, e que podem ajudar na concretização dos seus objetivos.

Evoluindo a arquitetura mostrada na Figura 3.3 uma arquitetura BDI é mostrada na Figura 3.4

3.2.3 Sistemas Multi-Agentes

Várias tendências da IA e da própria ciência da computação contribuíram para o desenvolvimento do conceito de um agente inteligente. Assim sendo, ainda é comum existirem, na literatura e textos académicos, explicações justificando algumas diferenças, tais como: o que é um agente; esclarecimentos sobre agentes ser realmente uma disciplina de IA; diferenciar agentes de sistemas especialistas; e a relação entre agentes e objetos. Wooldridge [129] procura oferecer um quadro definitivo sobre estas questões, deixando mais clara a posição e importância dos agentes inteligentes para a engenharia de software e para a IA.

3.2.3.1 Definição e Classificação

Para Issicaba [70] um sistema baseado em agentes significa aquele em que a abstração principal é a do agente. Portanto, um sistema de agente único refere-se a um sistema baseado em agente com apenas um agente, compreendendo então um ambiente de agente

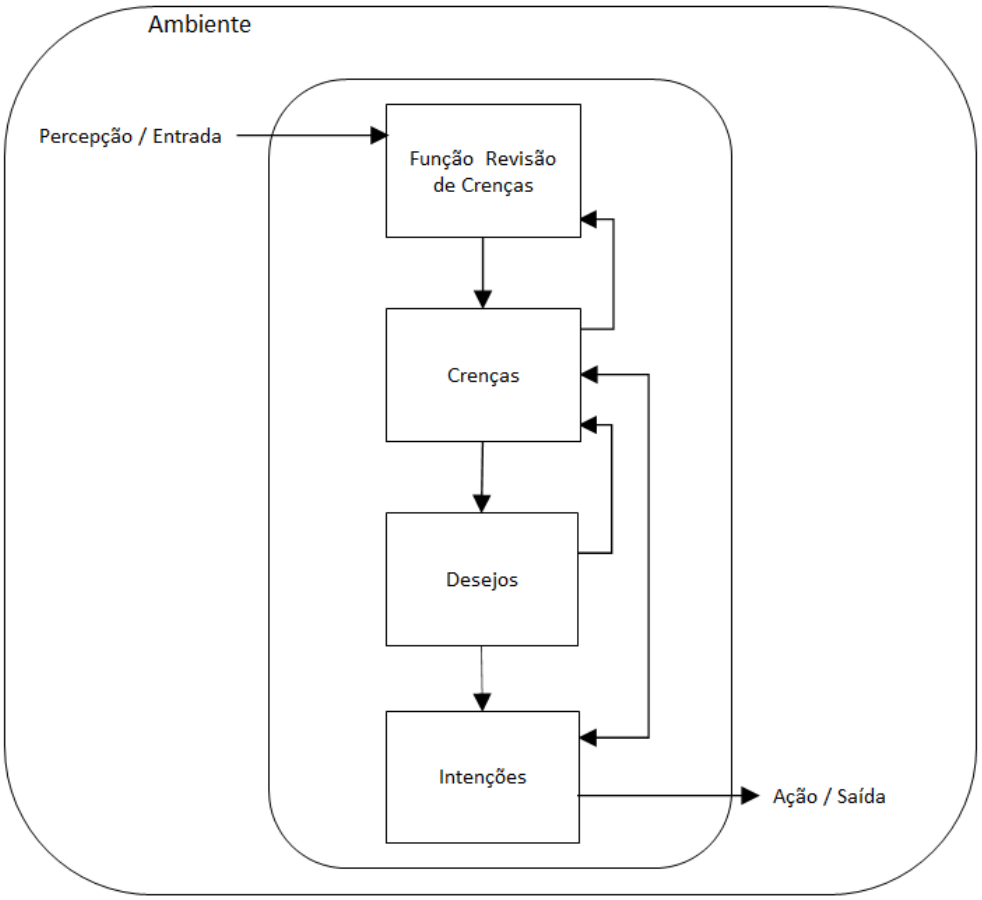


Figura 3.4: Arquitetura BDI de um Agente

único. Da mesma forma, um sistema multiagente refere-se a um sistema baseado em agente com mais de um agente, compreendendo então um ambiente multiagente. Cada agente possui conjuntos internos de estruturas e mecanismos que permitem que ele raciocine sobre si mesmo e sobre o ambiente.

A Resolução Distribuída de Problemas (RDP) desenvolve a forma como um problema pode ser resolvido por vários módulos, ou nós, que cooperam na resolução de um problema específico. Por outro lado, um SMA representa uma conjunto interligado de agentes trabalhando para resolver um problema cuja solução está para além da capacidade individual de um único agente. Num SMA não se pode assumir que todos os agentes possuem um objetivo comum, muitas vezes eles representam os interesses e objetivos de diferentes entidades ou organizações. Os interesses dos agentes podem gerar conflitos, tal como acontece nas sociedades humanas. Contudo, os agentes terão que cooperar de modo a atingir os seus objetivos, também à semelhança das sociedades humanas [39].

Uma taxonomia para os SMA é apresentada por Garcia and Sichman [66], nela os agentes são classificados segundo a perspetiva, abertura, granularidade, composição e interação [116]:

- Perspetiva - indica o objetivo do sistema. O sistema tem perspetiva de simulação,

quando seu objetivo são as interações sociais, e tem perspetiva de resolução, quando objetiva resolver problemas de forma cooperativa e distribuída;

- Abertura - o sistema pode ser aberto ou não, indicando a possibilidade do mesmo alterar sua composição dinamicamente, modificando o número de agentes que o compõe;

- Granularidade - indica se o sistema é composto por poucos agentes (baixa granularidade) ou muitos agentes (alta granularidade);

- Composição - indica se o sistema é formado por agentes homogéneos (agentes do mesmo tipo) ou heterogéneos;

- Interação - define os diversos tipos de relacionamentos entre os agentes componentes do sistema (comunicação, cooperação, coordenação e negociação).

3.2.3.2 Interação entre Agentes

Sistemas Multi-Agente (SMA) são sistemas distribuídos. Projetar e implementar um SMA significa especificar e codificar rigorosamente as comunicações entre os agentes por meio do protocolo de interação. Muitos consideram desafiador realizar esta especificação, considerando as propriedades dos agentes em serem autónomos e heterogéneos [32]. Esta comunicação entre agentes, também chamada de interação, é representada pela troca de informações, de forma direta (comunicação explícita) ou indireta (emissão de sinais através do ambiente).

Comprovadamente existe um esforço da comunidade em criar padrões de construção que permitam compartilhar e integrar agentes, quer eles tenham sido construídos ou não pelo mesmo grupo. Estes esforços são liderados pela *Foundation for Intelligent Physical Agents (FIPA)*[1] e *Knowledge Query and Manipulation Language (KQML)*[2]. A FIPA, criada em 1996, foi integrada ao *IEEE Computer Society (C/FIPA)* em 2005[3]. Em 2017 o *IEEE SA Standards Board* transferiu o C/FIPA para um projeto especial do *IEEE Computer Society Standards Activities Board (C/SAB Special Project the PAR - open project P2409)*[4].

A C/FIPA manteve integralmente as quatro linguagens propostas pela FIPA, inclusive com o nome original: *FIPA Semantic Language (FIPASL)*, *FIPA Constraint Choice Language (FIPA-CCL)*, *FIPA Knowledge Interchange Format (FIPA-KIF)* e *FIPA Resource Description Framework (FIPA-RDF)*.

[1] http://www.fipa.org/

[2] https://www.csee.umbc.edu/csee/research/kqml/papers/

[3] http://https://docplayer.net/5982064-Institute-of-electrical-and-electronics-engineers-ieee-fipa-standards-committee-fipa-sc-policies-and-procedures.html

[4] https://standards.ieee.org/about/sasb/resolutions/

A comunicação entre agentes precisa ser clara na intenção do ato comunicativo, por isso, as linguagens de comunicação entre agentes baseiam-se na teoria dos atos de fala[5]. Essa teoria apresenta uma forma de organizar a conversação, para que a mesma se torne mais eficaz. Cada mensagem deve conter, além do conteúdo, a indicação da intenção da comunicação, ou seja, sua tipologia, auxiliando o destinatário no entendimento e facilitando a obtenção dos objetivos. Essa mensagem pode ser enviada para um único agente (ponto-a-ponto) ou para vários agentes (*broadcast*).

Signoretti [116, p. 21] apresenta um bom resumo sobre o mecanismo de mensagens entre os agentes:

> "As mensagens trocadas pelos agentes são representadas usando duas linguagens: a externa e a interna. A linguagem interna é a linguagem usada para expressar o conteúdo da mensagem. Já a externa é usada para declarar os participantes da comunicação; a descrição do conteúdo (assunto) e o tipo de ato comunicativo (tipologia). Tanto a linguagem interna quanto a externa devem ser conhecidas por todos os participantes da comunicação. Como exemplo de linguagem externa tem-se a KQML e a *FIPA Agent Communications Language (FIPA-ACL)*, as linguagens internas usadas por elas são a FIPA-KIF e a *FIPA Semantic Language (FIPA-SL)*, respetivamente."

A FIPA-ACL continua sendo a principal referência para propostas de expansão da linguagem, como a apresentada por Rodriguez-Arias et al. [105] que utiliza os *communicative acts* da ACL no ambiente de jogos eletrónicos.

A FIPA também formalizou a Agent Unified Modelling Language (AUML) como diagrama de sequência para especificação do protocolo de interação[6].

Por vezes considerada linguagem e outras vezes como protocolo, a KQML permite a comunicação entre agentes e humanos. É uma linguagem estruturada em três camadas [66, 32]:

- Conteúdo - motivo da comunicação, podendo transportar qualquer linguagem de representação de conhecimento;
- Mensagem - determina as interações possíveis com um agente que se comunique em KQML, identificando o protocolo de entrega da mensagem e fornecendo uma ação ou ato de fala que o emissor anexa ao conteúdo. Este ato de fala identifica o conteúdo como uma consulta, uma assertiva ou um comando;
- Comunicação - refere-se aos parâmetros de identificação de emissor, destinatário, multiplicidade (única ou múltipla) do pacote ou quadro.

Urna mensagem gera um ato de fala, como *ask* ou *tell*. A semântica é informal, a linguagem é simples e independente de plataforma ou conteúdo. A sintaxe foi inspirada

[5] A teoria dos atos de fala foi elaborada inicialmente por John L. Austin (1911-1960) e desenvolvida posteriormente por J. R. Searle

[6] http://www.fipa.org/specs/fipa00025/XC00025E.html#_Toc505480202

na linguagem LISP[7]. A linguagem trata agentes e bases de conhecimento igualmente, para tanto os agentes terão uma interface, denominada *wrapper*, semelhante a das bases de conhecimento, que permita: perguntas e afirmações sobre seu conteúdo; inclusão ou exclusão de informações em seu conteúdo; pedidos ou ofertas de capacidades.

3.2.4 Agentes Aplicados ao Anti-Branqueamento de Capitais

Não é extensa a relação com as propostas que consideram o uso de abordagens baseadas em agente para suportar o Anti-Branqueamento de Capitais (ABC). Gao et al. [65], define uma arquitetura de sistema utilizando um conjunto de agentes especializados, tais como: agentes de coleta de dados (sistemas internos e informações externas); agentes de monitoramento para acompanhamento do perfil cadastral do cliente e das transações realizadas; e um agente que emite relatórios e alertas sobre possíveis operações de BC.

> *Apesar da boa proposta de arquitetura, o problema crucial do volume de análises submetidas ao analista humano não é enfrentando, aliás, pode até ser agravado com a automatização da fase de sinalização de transações suspeitas.*
>
> *(Nota L)*

Gao and Xu [64] propõem uma alteração na arquitetura apresentada por Gao et al. [65], incorporando as fases e técnicas de um modelo de tomada de decisão, aplicado em tempo real. O objetivo era melhorar as fases de identificação e sinalização de transações suspeitas do sistema multi-agente definido no trabalho anterior, na tarefa de ABC. O sistema utiliza três grupos de agentes: o *intelligence group*, com agentes responsáveis pela coleta de dados, perfilhamento de clientes e monitoramento de transações; o *design group*, onde uma categorização é realizada visando a realização de uma análise crítica por parte de agentes especializados; e o *choice group*, responsável pela escolha da melhor decisão, apresentação de relatórios e interface de usuário.

> *A arquitetura do sistema proposto anteriormente foi melhorada, mas em termos de fluxo não houve avanços, persistindo a falta de apoio aos analistas humanos.*
>
> *(Nota M)*

Outra abordagem baseada em agente foi apresentada por Xuan and Pengzhu [131]. A arquitetura proposta é semelhante à descrita por Gao et al. [65], porém, com evolução dos agentes incluindo as características de negociação, diagnóstico e autoaprendizagem. Um grupo de agentes chamados supervisores, em última análise, é responsável pelas decisões mais importantes, tomadas com base nas informações fornecidas pelo grupo de agentes de coleta de dados.

[7] http://www-formal.stanford.edu/jmc/lisp20th/lisp20th.html

As características dos agentes é o ponto forte desta proposta, contudo, também não detalha como melhorou a atuação dos analistas na passagem de sinalizações para o analista humano.

(Nota N)

Um sistema especialista que utiliza ontologia para detetar transações suspeitas é apresentado por Rajput et al. [103]. O sistema, que utilizou um bom volume de dados reais, consiste em um domínio de conhecimento e algumas regras para apoiar o raciocínio. A abordagem descrita utiliza as diretrizes de combate ao BC definidas pelo Banco do Paquistão. Um conjunto de regras definidas pelo autor descartam transações em função do tempo em que foram executadas, do limite de valor e de grupos estabelecidos.

O artigo não indica o volume de transações sinalizadas como suspeitas. Apesar do considerável volume de dados utilizados, as regras definidas pelos autores o universo analisado ficou muito limitado. Por exemplo, foram analisados somente transações entre 2.500 € e 12.000 € e que atendiam outras condições.

(Nota O)

3.2.5 Abordagem de Risco Aplicada ao Anti-Branqueamento de Capitais

Abordagem baseada em risco não é comumente utilizada no processo de Anti-Branqueamento de Capitais (ABC). Shijia Gao, um dos autores do trabalho [65], publicou, no mesmo ano, o trabalho [63] onde critica o trabalho anterior por dedicar grande foco nas regras de produção e propõe um modelo conceitual, ressaltando a busca no controle e prevenção. Uma das novidades da proposta é a utilização de indicadores de risco definidos com base em informações obtidas no cadastro do cliente.

Helmy et al. [69] apresenta um monitor para identificação de transações suspeitas baseado em dois mecanismos de risco: risco do cliente e risco da transação. Estes dois tipos de risco são aplicados em cenários pré-definidos e utilizam probabilidades para assinalarem as transações suspeitas. O risco do cliente é totalmente depende do Know Your Customer (KYC) e o risco da transação baseia-se em cenários pré-definidos para atribuição de pesos e aprimoramento do modelo de probabilidades.

Abordagens fortemente baseadas em informações cadastrais apresentam pouca eficiência pois não é possível garantir a precisão destas informações, nem que estejam atualizadas.

(Nota P)

Demetis and Angell [46] redefiniu o conceito de modelo de risco para Branqueamento de Capitais (BC) estendendo os modelos tradicionais com qualquer atributo que reduza a complexidade inicial do conjunto de transações. Nesse sentido, quaisquer dados de

fraude, marketing, demografia, media etc., podem ser usados. Demetis [45], também apresentou um estudo de caso de um banco do Reino Unido. O conceito estendido foi aplicado neste trabalho considerando que os atributos utilizados para definir os grupos de risco vão além das definições legais e preparam o ambiente para a incorporação de informações de outras fontes.

> *Apesar de expandir a possibilidade de atributos, os que foram utilizados no trabalho são de fontes cadastrais. Melhor em relação a propostas anteriores, mas ainda limitado ao Know Your Customer (KYC).*
>
> *(Nota Q)*

Uma proposta inovadora foi apresentada por Tai and Kan [117] onde os autores propõem um processo baseado em aprendizado por máquina e técnicas de análise de dados. É utilizado um modelo com 18 atributos, sendo oito baseados na média da quantidade dos tipos de transações e os demais na frequência de ocorrência dos dígitos. Numa segunda parte do modelo é baseada na Lei de Benford ou Lei do Primeiro Dígito que se refere a distribuição de dígitos nas fontes de dados.

de Jesús Rocha-Salazar et al. [40] também buscou ir além das informações obtidas nas políticas *Know Your Customer (KYC)* e propôs uma metodologia que considera variáveis não-transacionais, relacionadas com produtos e informações geográficas, variáveis presentes no relatório do *Financial Action Task Force (FATF)*. Estas variáveis servem de base para um indicador de anormalidade, que utiliza a variância destas variáveis.

> *Modelos baseados em variáveis quantitativas mostram aplicabilidade mais aderente ao Anti-Branqueamento de Capitais (ABC). A Lei de Benford tem apresentado mais eficácia em situações de fraude, em Branqueamento de Capitais (BC) a manipulação de números não é usual.*
>
> *(Nota R)*

3.2.6 Conclusão

Os conteúdos apresentados neste capítulo, notadamente das subsecções 3.1.4 - Aplicações para o Anti-Branqueamento de Capitais, 3.2.4 - Agentes Aplicados ao Anti-Branqueamento de Capitais e 3.2.5 - Abordagem de Risco Aplicada ao Anti-Branqueamento de Capitais, foram determinantes para a escolha da abordagem adotada no desenvolvimento deste trabalho, considerando que permitiu identificar a aderência dos trabalhos publicados com o tema Branqueamento de Capitais (BC) e evitou que abordagem já estudada e testada fosse adotada, sem que apresentasse contribuição para a melhoria do processo.

Os comentários apresentados, em forma de Notas, salientam os principais aspetos dos trabalhos publicados sobre o tema em estudo. Os aspetos positivos, considerados no desenvolvimento do trabalho, estão destacados nos outros capítulos deste trabalho, principalmente em Aplicando Mineração de Dados e Aprendizado de Máquina para ABC (Capítulo 4, pág. 45).

Capítulo 4

Aplicando Mineração de Dados e Aprendizado de Máquina para ABC

A exploração e utilização dos resultados encontrados na idealização da estratégia visando a solução do problema apresentado, delimitam os estudos realizados e que serão descritos neste capítulo. O objetivo é a formação de perfis dos clientes, oriundos, primordialmente, do histórico de utilização dos serviços disponibilizados pela instituição financeira. Estes perfis permitem o estabelecimento de um padrão de comportamento, tanto no uso dos serviços da instituição financeira, como na média dos valores transacionados. Este enfoque foi comentado em [79], na Nota D da secção 3.1.4 (pág. 32), na Nota J da secção 3.2.1 (pág. 34), além do modelo apresentado por Mak et al. [86].

Os dados utilizados na aprendizagem são reais e oriundos de uma instituição financeira brasileira cuja identificação, por questões de segurança e compromissos assumidos, será mantida em sigilo. Estes dados referem-se ao produto Contas-Correntes[1], o primeiro a ser implementado no sistema proposto.

4.1 Modelo e Descrição dos Dados

O conjunto de dados que compõe o produto Contas-Correntes é constituído de 6 tabelas, que originalmente possuem as seguintes descrições e interligações:

- **Cadastro** -– contém informações cadastrais dos clientes referentes ao sistema de contas-correntes. A cada cliente é atribuído um único código de identificação (Código de Cliente) e que será associado com todas as contas deste cliente no banco. Não existe limite para a quantidade de contas relacionadas a cada cliente. A identificação de cada uma das contas do cliente é formada pelo conjunto de informações:

[1] Conta-corrente é o principal produto oferecido pela banca pública e privada, também conhecida como conta à ordem, pode ser mantida por pessoa singular ou jurídica e permite a realização de transações em espécie ou eletrónicas. Estas transações podem envolver depósitos, saques, transferências, pagamentos, tanto a débito quanto a crédito, nacionais ou internacionais.

código da agência, número da conta-corrente e dígito verificador da conta. O código do cliente permite conexão com seus dados pessoais, não utilizados neste estudo;

- **Movimento** -– armazena informações sobre todas as transações realizadas pelos clientes. É organizada pelo dia de realização da transação e pela identificação da conta (código da agência, número da conta-corrente e dígito verificador da conta), permitindo relação com a tabela Cadastro. Armazena, também, os valores envolvidos e o saldo da conta após a conclusão da transação;

- **Histórico** -– apresenta um código que permite relação com a tabela Movimento e a descrição textual que detalha a natureza das transações previstas (tarifa, transferência de valores, depósito etc.). Pelo histórico é possível identificar, também, se a transações foi de crédito ou débito, em dinheiro, por cheque, online, realizado numa agência etc.;

- **Tipo de Pessoa** -– define a natureza contábil do cliente, indicando ser pessoa física ou singular, jurídica, governo, empresa nos diversos portes etc.;

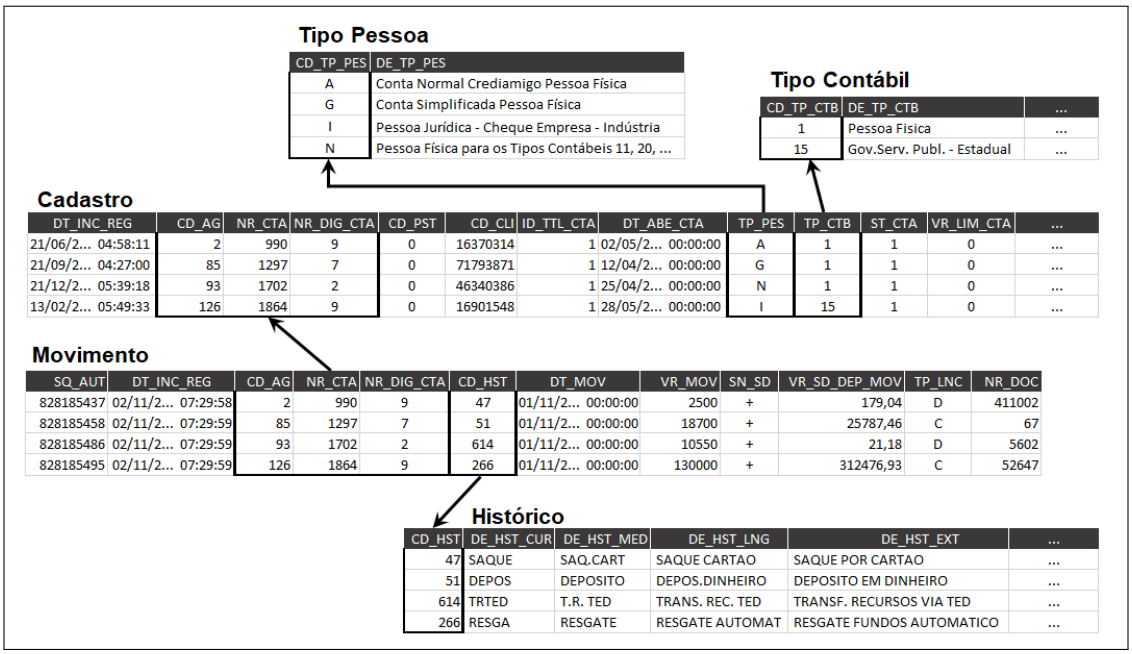
- **Tipo de Conta** -– determina o tipo de conta associada ao cliente. Podendo esta ser normal ou garantida, que é uma espécie de empréstimo rotativo etc.;

- **Tipo Contábil** -– relaciona a conta do cliente com um tipo contábil apropriado e que será utilizado pelo sistema de contabilidade. Assume valores como “Instituição Financeira Cooperativa”, “Governo Administração Direta Federal” etc.

A descrição dos campos existentes em cada tabela, bem como o diagrama mostrando o relacionamento entre as tabelas pode ser visto no Apêndice A, contudo, a Figura 4.1 mostra um exemplo dos dados contidos em cada tabela. A Figura 4.2 apresenta um resumo do modelo dos dados envolvidos no processo, este resumo será referenciado em fluxos que serão apresentados nas secções seguintes.

Considerando o grande volume de dados e por ser a primeira vez que esta base passaria por este tipo de análise, a opção foi por analisar inicialmente apenas três meses de movimentação. Outro fator que contribuiu para utilização de um conjunto reduzido de dados, foi a curva de aprendizado do autor do trabalho sobre as ferramentas a serem utilizadas, resultando em repetidas execuções de rotinas, por falhas na conceção ou busca de melhoria, o que seria inviável com um grande volume de dados.

Após o mapeamento inicial e o devido entendimento das bases de dados e das ferramentas utilizadas, todas as transações referentes a um ano de movimentação passaram a ser utilizadas, ou seja, foram analisadas 90,6 milhões de transações, efetuadas por alguns dos 5,1 milhões de clientes. Os volumes envolvidos são detalhados na Tabela 4.1. Os dados foram armazenados no banco de dados *MySQL Workbench (MySQL)* [91], considerando que este BD permite integração com as ferramentas a serem utilizadas para mineração dos dados.



Tipo Pessoa

CD_TP_PES	DE_TP_PES
A	Conta Normal Crediamigo Pessoa Física
G	Conta Simplificada Pessoa Física
I	Pessoa Jurídica - Cheque Empresa - Indústria
N	Pessoa Física para os Tipos Contábeis 11, 20, ...

Tipo Contábil

CD_TP_CTB	DE_TP_CTB	...
1	Pessoa Fisica	...
15	Gov.Serv. Publ. - Estadual	...

Cadastro

DT_INC_REG	CD_AG	NR_CTA	NR_DIG_CTA	CD_PST	CD_CLI	ID_TTL_CTA	DT_ABE_CTA	TP_PES	TP_CTB	ST_CTA	VR_LIM_CTA	...
21/06/2... 04:58:11	2	990	9	0	16370314	1	02/05/2... 00:00:00	A	1	1	0	...
21/09/2... 04:27:00	85	1297	7	0	71793871	1	12/04/2... 00:00:00	G	1	1	0	...
21/12/2... 05:39:18	93	1702	2	0	46340386	1	25/04/2... 00:00:00	N	1	1	0	...
13/02/2... 05:49:33	126	1864	9	0	16901548	1	28/05/2... 00:00:00	I	15	1	0	...

Movimento

SQ_AUT	DT_INC_REG	CD_AG	NR_CTA	NR_DIG_CTA	CD_HST	DT_MOV	VR_MOV	SN_SD	VR_SD_DEP_MOV	TP_LNC	NR_DOC
828185437	02/11/2... 07:29:58	2	990	9	47	01/11/2... 00:00:00	2500	+	179,04	D	411002
828185458	02/11/2... 07:29:59	85	1297	7	51	01/11/2... 00:00:00	18700	+	25787,46	C	67
828185486	02/11/2... 07:29:59	93	1702	2	614	01/11/2... 00:00:00	10550	+	21,18	D	5602
828185495	02/11/2... 07:29:59	126	1864	9	266	01/11/2... 00:00:00	130000	+	312476,93	C	52647

Histórico

CD_HST	DE_HST_CUR	DE_HST_MED	DE_HST_LNG	DE_HST_EXT	...
47	SAQUE	SAQ.CART	SAQUE CARTAO	SAQUE POR CARTAO	...
51	DEPOS	DEPOSITO	DEPOS.DINHEIRO	DEPOSITO EM DINHEIRO	...
614	TRTED	T.R. TED	TRANS. REC. TED	TRANSF. RECURSOS VIA TED	...
266	RESGA	RESGATE	RESGATE AUTOMAT	RESGATE FUNDOS AUTOMATICO	...

Figura 4.1: Exemplo dos Dados nas Tabelas

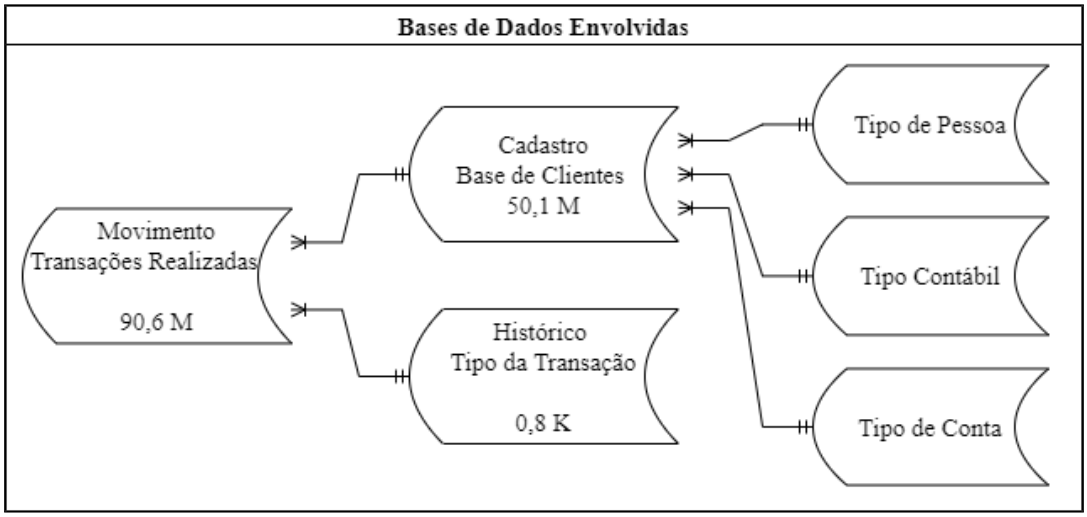


Figura 4.2: Resumo do Modelo dos Dados Envolvidos no Processo

4.2 Técnicas e Ferramentas Utilizadas

O combate ao Branqueamento de Capitais (BC) caracteriza-se por ser um processo de descoberta de conhecimento, por vezes escondido no meio de bases de dados gigantescas, assim sendo, a mineração de dados torna-se essencial. A análise de transações e de informações sobre as contas, com agrupamento, classificação e mineração baseada em restrições pode melhorar a eficácia, a precisão e, também, descobrir novos padrões de BC [83].

Os métodos supervisionados não são adequados para a maioria das instituições finan-

Tabela 4.1: Volumetria das Bases de Dados Utilizadas

Nome da Tabela	Quantidade de Linhas	Tamanho da Tabela (MB)
Movimento	90.630.515	8.999.702
Cadastro	5.113.396	545.138
Histórico	838	0.096
Tipo Contábil	59	0,004
Tipo de Pessoa	37	0.004
Tipo de Conta	2	0.002

ceiras porque exigem dados reais e, geralmente, o número de casos de BC confirmados como suspeitos é pequeno se comparado com a quantidade de casos sinalizados e analisados e para os quais existe uma justificativa aceitável. Esta situação dificulta a construção de um conjunto de dados rotulado e a aplicação de algoritmos supervisionados. Dentre os algoritmos não-supervisionados, os de agrupamento têm sido os mais utilizados neste domínio [40].

Como será explicado na secção seguinte, os dados utilizados neste trabalho não possuem uma classe (não-supervisionado). O objetivo inicial é descobrir padrões a partir de alguma característica de regularidade presente no conjunto de dados (*clustering*), buscando a formação de grupos de clientes com características semelhantes e mutuamente exclusivos (particionado). A expetativa era que as características comuns dos clientes permitisse a identificação de grupos com diferentes escalas de suspeições para possíveis atividades de BC, podendo vir a ser classificado como grupos de risco. A Figura 4.3, adaptada de [88], mostra a hierarquia de aprendizagem com o caminho escolhido neste trabalho.

Vários estudos já foram realizados com conclusões semelhantes sobre os algoritmos particionados serem mais eficientes ao lidar com grandes volumes de dados. Os algoritmos hierárquicos são muito mais caros computacionalmente do que K-means, por exemplo, o que pode ser uma barreira para usá-los em conjuntos de dados muito grandes [73, 71, 58].

O algoritmo K-means [67] foi o escolhido, apesar de repetidamente ser salientado como obstáculo à sua utilização a necessidade de se definir a priori o número de *clusters* a ser utilizado, contudo, ele é um dos métodos mais utilizados. Esta popularidade talvez possa ser explicada por ele possibilitar bons resultados, pela sua simplicidade, ser de fácil entendimento, eficiência e por estar presente em quase todos os ambientes de implementação e automatização do processo de *data mining* [108, 86].

Outra questão que precisa ser tratada quando da utilização do K-means é quanto ao fato dos melhores resultados serem apresentados com atributos numéricos contínuos em comparação com resultados utilizando atributos nominais. A explicação reside no fato de, originalmente, ser utilizado o quadrado da distância Euclidiana para cálculo de proximidade. Em atributos nominais este cálculo não pode ser realizado [98, 52].

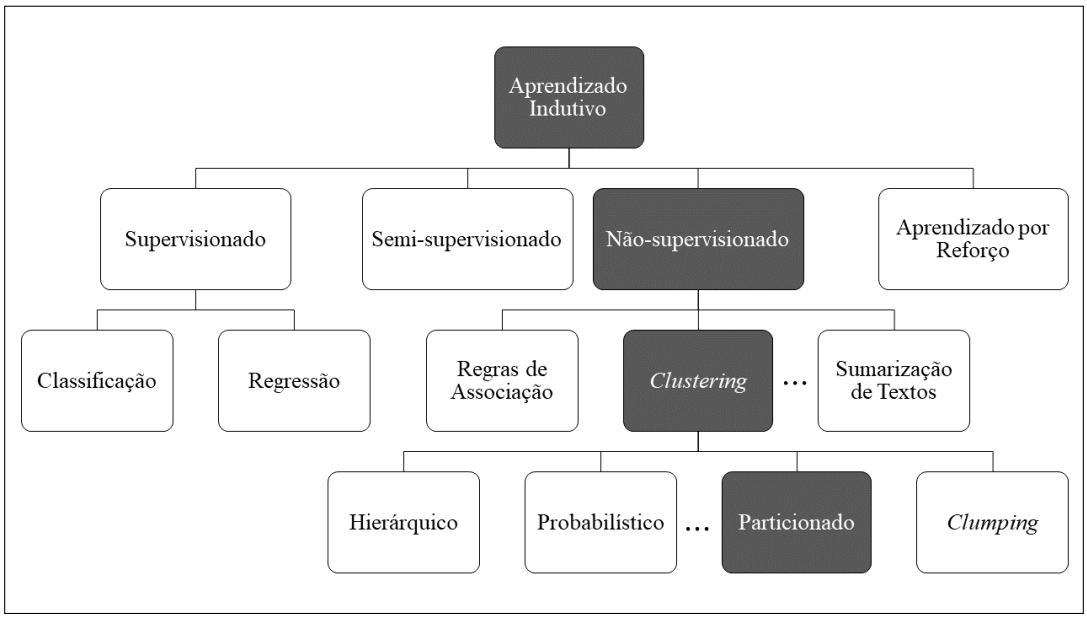


Figura 4.3: Hierarquia do Aprendizado Indutivo

A facilidade de uso e a eficiência documentada da ferramenta *Waikato Environment for Knowledge Analysis (WEKA)*[2], produto da Universidade de Waikato[3] (Nova Zelândia), determinou a escolha deste ambiente, aliada ao fato desta possuir suporte para todas as etapas do processo de Mineração de Dados (MD), com uma boa interface gráfica além de implementar vários algoritmos de *clustering*, de forma nativa, incluindo o *SimpleKMeans* [9].

Esta versão do K-means no WEKA aplica, tanto para distância Euclidiana como para distância *Manhattan*, a técnica do vizinho mais próximo visando reduzir o problema com os atributos nominais. A regra geral é que para dois valores de atributos numéricos X e Y, o resultado de X-Y é utilizado no cálculo da distância. Quando o atributo é nominal é atribuído o valor 0 quando X e Y são iguais e 1 quando são diferentes.

Também foi utilizada a ferramenta *Massive Online Analysis (MOA)*[4], da Universidade de Waikato. Enquanto a WEKA facilita o manuseio, limpeza e tratamentos dos dados, a MOA possibilita a realização de avaliações dos *clusters* gerados, pois ela incorpora mais algoritmos que a Weka e utiliza uma forma diferente de avaliação, baseada na divisão da base de dados para treinamento e teste [68, 19].

Na busca pelo número adequado de *clusters* a ser utilizado, o algoritmo *SimpleKMeans* foi executado 10 vezes e as métricas mais populares, tais como Coeficiente Silhueta [106], *Sum of Squared Error (SSE)*, *Variance Ratio Criterion (VRC)* [26], Van Dongen e Rand [123] foram analisadas, utilizando MOA. Além de comparação realizada com o resultado de execução do algoritmo CascadeSimpleKMeans [26], que se propõe a indicar

[2]http://www.cs.waikato.ac.nz/ml/weka/index.html
[3]http://www.cs.waikato.ac.nz/
[4]http://moa.cms.waikato.ac.nz/

o melhor número de *clusters*.

Com os *clusters* gerados, a etapa seguinte é da geração de regras. Para cumpri-la foram realizados experimentos com os seguintes algoritmos, todos incorporados ao ambiente WEKA: PART [61], que implementa o algoritmo C4.5 *Decision Tree* para as interações e utiliza a técnica *best leaf* na geração das regras; J48 [101] também baseado no C4.5, mas com a opção de "*reducedErrorPruning*"; e JRip [37], que visa maximizar o resultado da técnica de poda sucessiva.

4.3 Tratamento dos Dados

Conforme já mencionado, num primeiro momento o estudo foi realizado com apenas três meses de movimentação, algo em torno de 14,5 milhões de linhas de transações, além de 4,5 milhões de linhas referente a dados cadastrais. Na etapa de pré-processamento dos dados, seguindo procedimento consagrado, buscou-se a melhoria da qualidade dos dados [62]. Os dados foram então agrupados por cliente com atributos numéricos que indicavam:

1. Tipo de pessoa;
2. Idade da Conta;
3. Média da quantidade de serviços utilizados no período;
4. Média da quantidade total de transações realizadas no período;
5. Média da quantidade de transações realizadas a crédito no período;
6. Média da quantidade de transações realizadas a débito no período;
7. Valor médio das transações a crédito realizadas no período;
8. Valor médio das transações a débito realizadas no período;
9. Desvio padrão da quantidade serviços utilizados no período;
10. Desvio padrão da quantidade total de transações realizadas no período;
11. Desvio padrão da quantidade de transações realizadas a crédito no período;
12. Desvio padrão da quantidade de transações realizadas a débito no período;
13. Desvio padrão do valor das transações a crédito no período;
14. Desvio padrão do valor das transações a débito no período;
15. Coeficiente de variação do valor de transações a crédito no período;
16. Coeficiente de variação do valor de transações a débito no período;

Uma tabela com 1,6 milhões de linhas foi o resultado desta forma de organização dos clientes.

Apesar das já citadas restrições quanto a utilização de atributos nominais na formação de *clusters* com o K-means e seus similares, para cada um dos atributos numéricos foi efetuada uma discretização com quatro ou cinco faixas de valores, dependendo da amplitude dos valores. Três cenários foram testados: com todos os atributos numéricos; com todos os atributos nominais; com atributos numéricos e nominais. A motivação para esta decisão foi a grande variação entre os valores mínimo e máximo dos atributos, onde a maior variação ficou entre o mínimo 0,01 e o máximo 536.852.446,89.

Apesar dos *clusters* gerados terem apresentado coerência com a realidade dos clientes, análise corroborada por analistas do banco fornecedor dos dados, o resultado final alcançado não pode ser considerado satisfatório, posto que as regras obtidas não permitiam qualificar os clientes por vários atributos. Muitas regras utilizavam somente um atributo, selecionando uma grande quantidade de clientes que não representavam nenhum risco de BC. Por exemplo, regras que apenas separavam os clientes por tipo de pessoa (singular ou jurídica). Assim sendo, uma nova estratégia foi adotada para o trabalho a ser desenvolvido com o volume de dados citado na Tabela 4.1, referente a 1 ano de movimentação [2].

4.3.1 Definição do Perfil do Cliente

A nova estratégia foi baseada numa outra forma de apresentar as informações do perfil dos clientes, cuja estrutura é apresentada a seguir:

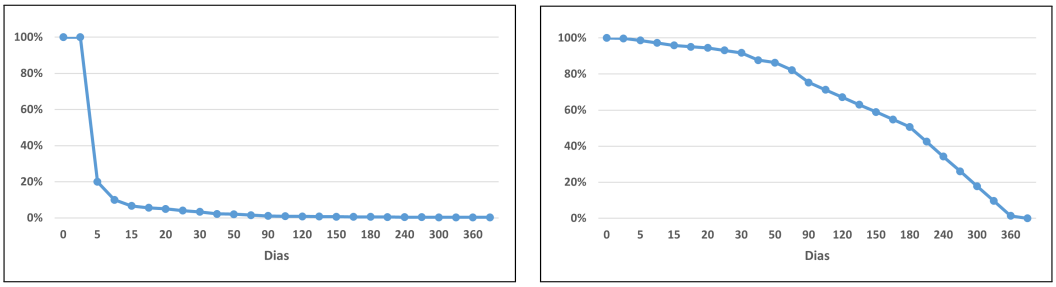
1. Código do Cliente;
2. Código da Agência;
3. Número da Conta Corrente;
4. Número do Dígito Verificador da Conta;
5. Tipo de Pessoa;
6. Tipo Contábil;
7. Conta Especial (S/N);
8. Idade da Conta (em anos);
9. Quantidade de Serviços Utilizados no Período;
10. Quantidade de Movimentos Realizados no Período;
11. Quantidade de Movimentos na faixa de valor [0,00 – 500,00];
12. Quantidade de Movimentos na faixa de valor (500,00 – 1.500,00];
13. Quantidade de Movimentos na faixa de valor (1.500,00 – 5.000,00];
14. Quantidade de Movimentos na faixa de valor (5.000,00 – 10.000,00];
15. Quantidade de Movimentos na faixa de valor (10.000,00 – 90.000,00];
16. Quantidade de Movimentos na faixa de valor (90.000,00 – inf];
17. Percentual do Montante de Débitos sobre o Montante de Créditos;
18. Percentual da Quantidade de Transferência Eletrónica Disponível (TED)s sobre a Quantidade de Créditos;
19. Percentual da Quantidade de Documento de Ordem de Crédito (DOC)s sobre a Quantidade de Créditos.

Os atributos 5 e 8 a 19 foram utilizados no aprendizado para formação de *clusters*. O atributo 5 é originalmente nominal e os demais são numéricos, sendo uma estrutura mais adequada para utilização do algoritmo *SimpleKMeans*. Os atributos 11 a 16 expressam contagem para faixa de valores em Reais, moeda brasileira de onde a base de dados é oriunda. DOC e TED identificam transações realizadas entre bancos e destinam-se a

transferência de recursos, informações cruciais para o presente estudo. A TED realiza a transferência de valores de forma imediata, enquanto o DOC é realizado somente no processamento *batch* noturno. Neste caso, a TED é uma transação de altíssimo risco e preferencial nos casos de Branqueamento de Capitais (BC), por não possibilitar cancelamento.

Uma das características comportamentais do BC é a rapidez com que o dinheiro entra e sai da conta utilizada para o crime, geralmente esta conta recebe o nome de conta de passagem. O objetivo do fraudador é dificultar o rastreamento do dinheiro sujo, por isso ele utiliza várias contas em vários bancos para receber e repassar o dinheiro. O atributo 17 congrega duas informações importantes na busca de identificação desta característica no perfil do cliente: a percentagem de débitos sobre os créditos realizados no período; e o tempo decorrido entre a entrada e a saída do dinheiro da conta deste cliente. Para obter este atributo a referida percentagem foi ponderada pelo tempo de permanência destes valores na instituição financeira. Este atributo indica a volatilidade do dinheiro na conta do cliente, esta volatilidade é diretamente proporcional a um possível risco deste perfil. Quanto mais rápido o dinheiro é retirado da conta, mais suspeitas são essas transações.

Para identificar o comportamento foi analisado o cenário em que o cliente retira todo o montante depositado, ponderado pelo número de dias que permaneceu na conta. Esta ponderação produz uma curva com queda rápida nos primeiros dias e depois uma certa estabilidade na curva, praticamente a partir do trigésimo dia (Figura 4.4(a)). Com este comportamento o atributo não contribuiria para análise. Em outro cenário analisado uma ponderação foi realizada, também, pelos dias que o dinheiro permanece na conta, porém, proporcional a um ano de movimentação. O resultado foi uma curva com queda menos acentuada, mostrando-se mais adequada para o propósito deste trabalho (Figura 4.4(b)). A interpretação geral é: quanto maior o percentual de volatilidade maior o risco de tratar-se de um com transações suspeitas.



(a) Volatilidade com ponderação em dias

(b) Volatilidade com ponderação proporcional a um ano

Figura 4.4: Comparativo das Curvas da Ponderação do Atributo de Volatilidade

A Equação 4.1 implementa o conceito de volatilidade deste atributo [3].

$$PD(C) = \sum_{k=1}^{n_C} \frac{vd_k(C) * (1 - ((dd_k(C) - dc_1(C))/365)}{vc_k(C)} * 100 \tag{4.1}$$
$$para \quad vc_k(C) > 0, \quad vd_k(C) > 0 \quad e \quad dd_k \geqslant dc_1$$

onde:

PD(C) é a percentagem de débito de um cliente *C* específico; n_C é o número de registos para o cliente *C* na tabela de perfis; vd_k é o montante da k-ésima transação de débito realizada no período analisado; dd_k é a data da k-ésima transação de débito realizada no período analisado; vc_k é o montante da k-ésima transação de crédito realizada no período analisado; dc_1 é a data da primeira transação de crédito realizada no período analisado. A diferença entre duas datas é medida em dias, ponderado para um período de um ano.

4.3.2 Geração dos *Clusters* e Regras - Abordagem 1

Analistas de ABC da instituição financeira fornecedora da base de dados consultados, além de ratificarem a importância dos atributos definidos, identificaram os tipos de transações que não têm nenhuma relação com o crime de BC. Por exemplo, cobrança de tarifas realizadas pelo banco. Com esta definição a quantidade de linhas da tabela de perfis foi reduzida e a expectativa era que os *clusters* gerados seriam mais especializados em transações realmente passíveis de serem utilizadas no BC. A nova tabela de perfis de clientes resultou em 2,4 milhões de linhas.

O resultado da execução do algoritmo na busca pelo número adequado de *clusters* a ser utilizado pode ser verificado na Figura 4.5.

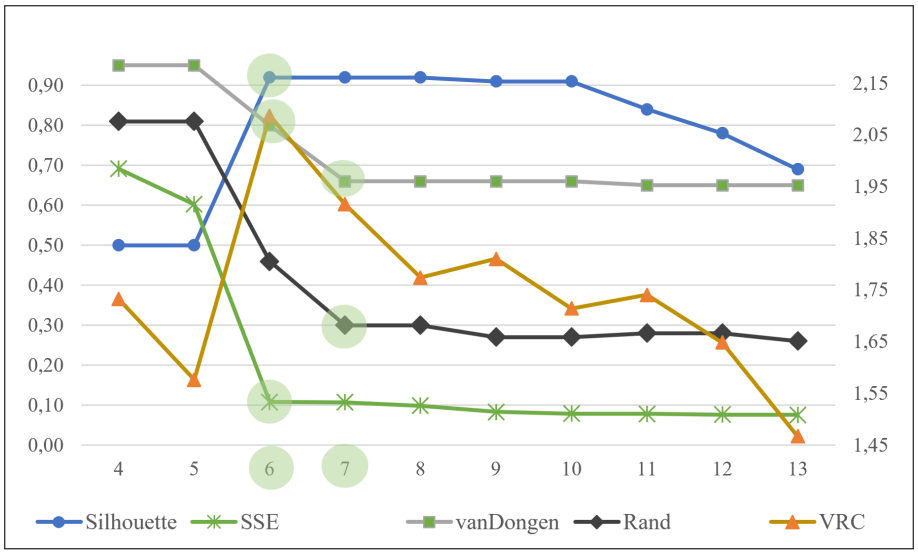


Figura 4.5: Busca do Número Ideal de *Clusters*

As métricas Silhueta, SSE e VRC indicam 6 *clusters* como o número ideal, enquanto

vanDongen e Rand indicam 7 *clusters*. Tanto Silhueta como SSE mostram uma linha de estabilidade a partir do número 6, o que pode corroborar o número 7 sinalizado pelas demais métricas. Contudo, o procedimento padrão é fazer a escolha recair sobre o "cotovelo" da curva ou o valor mais alto, dependendo da métrica. Portanto, inicialmente foram considerados 6 *clusters*.

Tanto os *clusters* gerados quanto os relatórios do ambiente de avaliação (WEKA), iniciam a numeração dos *clusters* em zero, contudo, na explicação abaixo e ao longo do texto a numeração será iniciada em um. As descrições abaixo são resultado de análise realizada para identificação das características comuns, estas descrições foram validadas pelos Analistas de ABC da instituição financeira:

- ***Cluster*** #1 – **Baixa Utilização**: grupo de clientes com baixíssima utilização de serviços e pouca movimentação. Valores manuseados também muito baixos e baixo percentual de débitos sobre os créditos realizados. Pode significar contas praticamente inativas com raras movimentações;

- ***Cluster*** #2 – **Alta Utilização**: clientes com maior volume de utilização de serviços e grande movimentação. Valores manuseados distribuídos por todas as faixas, com predominância nas faixas menores, porém, são os maiores usuários das faixas que representam grandes valores. Percentual alto de débitos sobre os créditos realizados com moderada utilização de TED (2ª maior utilização). Pode representar os grandes clientes do Banco;

- ***Cluster*** #3 – **Cliente Padrão**: maior grupo de clientes com utilização de serviços, movimentação e valores transacionados bem acima do "*cluster* #1" mas ainda bem abaixo do "*cluster* #2", ou seja, clientes intermediários. Alto percentual de débitos sobre os créditos realizados, significando que o dinheiro entra e sai da conta rapidamente. Utilização "normal" de TED. Pode representar a média dos clientes do Banco;

- ***Cluster*** #4 – **Grupo de Risco 1**: terceira maior movimentação em quantidade, com utilização de poucos serviços. Valores transacionados concentrados nas faixas mais baixas. Maior percentual de débitos sobre os créditos realizados e maior percentual de utilização de TED, significando que o dinheiro entra e sai da conta rapidamente via TED. Esta utilização de TED aliada à maior utilização de DOC torna este um grupo de alto risco, considerando também que os valores manuseados estão divididos em faixas menores;

- ***Cluster*** #5 – **Clientes Tradicionais**: clientes com maior tempo de abertura de contas, utilização de variada quantidade de serviços e "boa" quantidade de movimentos. Valores manuseados distribuídos por quase todas as faixas, com menor incidência na faixa mais alta. Pouco acima da metade dos créditos realizados permanecem no banco e com moderada utilização de TED;

- ***Cluster*** #6 – **Grupo de Risco Moderado**: grupo semelhante ao do "*cluster* #3" porém com mais volume financeiro envolvido, próximo ao limite legal. "Boa" utilização de serviços e quantidade de movimentos. Altíssimo percentual de débitos sobre os créditos realizados. Utilização de TED acima da média. Apesar da semelhança com o grupo do "*cluster* #3", considerando a quantidade de clientes e a maior utilização de TED torna este um grupo de risco moderado.

A pouca diferença entre os valores 6 e 7 para a quantidade total sugerida de *clusters* sinaliza a necessidade de verificação do resultado com 7 *clusters*. A redistribuição das instâncias para criação do sétimo *cluster* não afetou as características básicas dos seis *clusters* iniciais. Em termos proporcionais o *Cluster* #3 foi o que mais cedeu elementos para a formação do *Cluster* #7, como mostra a Tabela 4.2 e cujas características podem ser assim definidas:

- ***Cluster*** #7 – Grupo de Risco 2: perfil de contas mais antigas com grande utilização de serviços e grande volume de transações. Valores financeiros concentrados nas faixas denominadas de "limites legais". Maior percentual de saída de recursos financeiros, porém com baixa transferência para outras instituições. Alta transferências entre contas da mesma instituição.

As características apresentadas pelo sétimo *Cluster* torna importante a sua manutenção na configuração do sistema, dessa forma, o trabalho passou a utilizar a formação de 7 *clusters* no tratamento integral da base de dados.

Tabela 4.2: Comparação entre as Instâncias Clusterizadas

Treinamento (66%)		Teste (34%)		Treinamento (66%)		Teste (34%)	
#Cluster / Instâncias				#Cluster / Instâncias			
1	156.371 (10%)	1	80.456 (10%)	1	156.226 (10%)	1	80.386 (10%)
2	60.523 (4%)	2	31.083 (4%)	2	57.773 (4%)	2	29.676 (4%)
3	709.927 (45%)	3	366.331 (45%)	3	650.410 (41%)	3	335.588 (41%)
4	31.745 (2%)	4	16.227 (2%)	4	31.771 (2%)	4	16.247 (2%)
5	113.155 (7%)	5	57.604 (7%)	5	113.164 (7%)	5	57.593 (7%)
6	519.701 (33%)	6	268.122 (33%)	6	517.392 (33%)	6	266.943 (33%)
	1.591.422 (100%)		819.823 (100%)	7	64.686 (4%)	7	33.390 (4%)
					1.591.422 (100%)		819.823 (100%)

Conforme mencionado na secção 4.2 (pág. 47) as ferramentas WEKA e MOA permitem avaliar os *clusters* gerados através da utilização de várias medidas. Por padrão estas ferramentas consideram o último atributo como a classe dos dados, por isso, os *clusters* gerados foram incorporados à tabela de perfis.

As avaliações realizadas foram todas satisfatórias[5], a matriz de confusão, por exemplo, mostrou um nível de acerto acima de 99% considerando-se as taxas de classificação

[5]Todas as execuções para avaliação dos *clusters* foram realizadas utilizando a função de *Split* da base de dados na proporção 66% para treinamento e 34% para teste.

incorreta de 0,0683% e 0,0596%, respetivamente para a base de treinamento e para a base de teste, conforme mostram as Figuras 4.6 e 4.7.

```
=== Evaluation result for training instances ===
Scheme: SimpleKMeans
Relation: QueryResult-weka.filters.unsupervised.attribute.AddCluster-Wweka.clusterers.SimpleKMeans
-init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 7 -A
"weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 8-I1,2,3,4,6,7-
weka.filters.unsupervised.attribute.Remove-Rfirst-4,6,7

kMeans
======
Number of iterations: 22
Within cluster sum of squared errors: 106330.06201957827
Clustered Instances
0       156224 ( 10%)
1        57158 (  4%)
2       650723 ( 41%)
3        31797 (  2%)
4       113144 (  7%)
5       517423 ( 33%)
6        64953 (  4%)

Class attribute: perfil
Classes to Clusters:
      0      1      2      3      4      5      6  <-- assigned to cluster
 156224      0      0      0      2      0      0 | cluster1
      0  57152     10     44     25     40    556 | cluster2
      0      3 650380      0      0      1     33 | cluster3
      0      0      0  31753      0      0      2 | cluster4
      0      2      0      0 113111      0      0 | cluster5
      0      1     13      0      6 517354      1 | cluster6
      0      0    320      0      0     28  64361 | cluster7

Incorrectly clustered instances :     1087.0       0.0683 %
```

Figura 4.6: Avaliação dos *Clusters* (instâncias de treinamento)

A Figura 4.8 apresenta um fluxo com o objetivo de resumir as etapas executadas numa primeira abordagem. O elemento "bases de dados envolvidas" se refere ao modelo apresentado na Figura 4.2 (pág. 47). Os processos Seleciona Transações Relevantes, Estrutura Perfis de Clientes e gera *clusters* com o algoritmo K-Means, foram concluídos.

Seguindo o objetivo do trabalho agora deve se buscar as regras que melhor delimitem e identifiquem os *clusters* existentes. Estas regras possibilitarão a classificação de novos clientes e reclassificação de clientes que apresentem comportamento diferente daquele utilizado no aprendizado. A busca por estas melhores regras levou a realização de experimentos com os já citados algoritmos PART, J48 e JRip, utilizando-se os parâmetros *default* da ferramenta *WEKA* e restringindo a cobertura das regras a 100 e 1000 instâncias. Esta configuração (15 experimentos) foi executada 2 vezes: uma realizando *Split* da base de dados na proporção 66% para treinamento e 34% para teste; e outra com *cross-validation* (10 *folds*). Ou seja, no total foram realizadas 30 execuções, cujo resultado é mostrado na Figura 4.9.

O algoritmo JRip, apesar de gerar uma quantidade pequena de regras, o que seria ideal para trabalhar, no presente estudo, apresentou sempre baixa qualidade, conforme as métricas utilizadas. O algoritmo PART apresentou o melhor resultado, nos dois conjuntos de experimentos, para as variáveis analisadas, quando o número mínimo de instâncias

```
=== Evaluation result for test instances ===
Scheme: SimpleKMeans
Relation: QueryResult-weka.filters.unsupervised.attribute.AddCluster-Wweka.clusterers.SimpleKMeans
-init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 7 -A
"weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 8-I1,2,3,4,6,7-
weka.filters.unsupervised.attribute.Remove-Rfirst-4,6,7

kMeans
======
Number of iterations: 22
Within cluster sum of squared errors: 106330.06201957827
Clustered Instances
0       80386 ( 10%)
1       29318 (  4%)
2      335722 ( 41%)
3       16289 (  2%)
4       57652 (  7%)
5      266982 ( 33%)
6       33474 (  4%)

Class attribute: perfil
Classes to Clusters:
      0      1      2      3      4      5      6  <-- assigned to cluster
  80386      0      0      0      0      0      0 | cluster1
      0  29312      6     28     11     20    245 | cluster2
      0      1 335569      0      0      1     10 | cluster3
      0      2      0  16261      0      0      0 | cluster4
      0      3      0      0  57640      1      0 | cluster5
      0      0      8      0      1 266949      2 | cluster6
      0      0    139      0      0     11  33217 | cluster7

Incorrectly clustered instances :     489.0   0.0596 %
```

Figura 4.7: Avaliação dos *Clusters* (instâncias de teste)

por regra é limitado a 100. Contudo, o percentual de 71,72% de classificação correta, em princípio, não é um bom percentual. As regras geradas também não apresentaram a qualidade esperada, com estranhas repetições de atributos ou conflito nos atributos.

A consideração de que estes resultados não foram satisfatórios, motivou a realização da discretização de alguns atributos dos perfis. Dos atributos descritos na subsecção 4.3.1 Definição do Perfil do Cliente (pág. 51) foram discretizados os atributos 8 a 19, utilizando três grupos de valores com percentuais equivalentes de quantidade de ocorrências, quando possível. A Tabela 4.3 mostra um exemplo desta discretização.

Alguns atributos, devido a uma concentração de ocorrências em um único valor, foram divididos em apenas dois grupos (atributos 15, 16, 18 e 19). Os experimentos foram novamente realizados executando-se os mesmos algoritmos e utilizando-se os mesmos parâmetros do experimento anterior e os resultados são mostrados na Figura 4.10.

O algoritmo JRip manteve a mesma baixa qualidade nos resultados e por isso suas regras foram descartadas. Os algoritmos J48 e PART apresentaram os melhores resultados no conjunto de experimentos utilizando *Split* e *cross-validation*, respetivamente. Nos dois casos o número 1.000 limitou o mínimo de instâncias por regra para o PART e o mínimo de folhas na árvore do J48. Apesar de não ter sido significativa os indicadores apresentaram melhores resultados se comparados com os experimentos anteriores.

Ampliando mais o experimento com o propósito de verificar se o comportamento desses indicadores sofreria grande impacto com a variação na poda, os algoritmos PART e

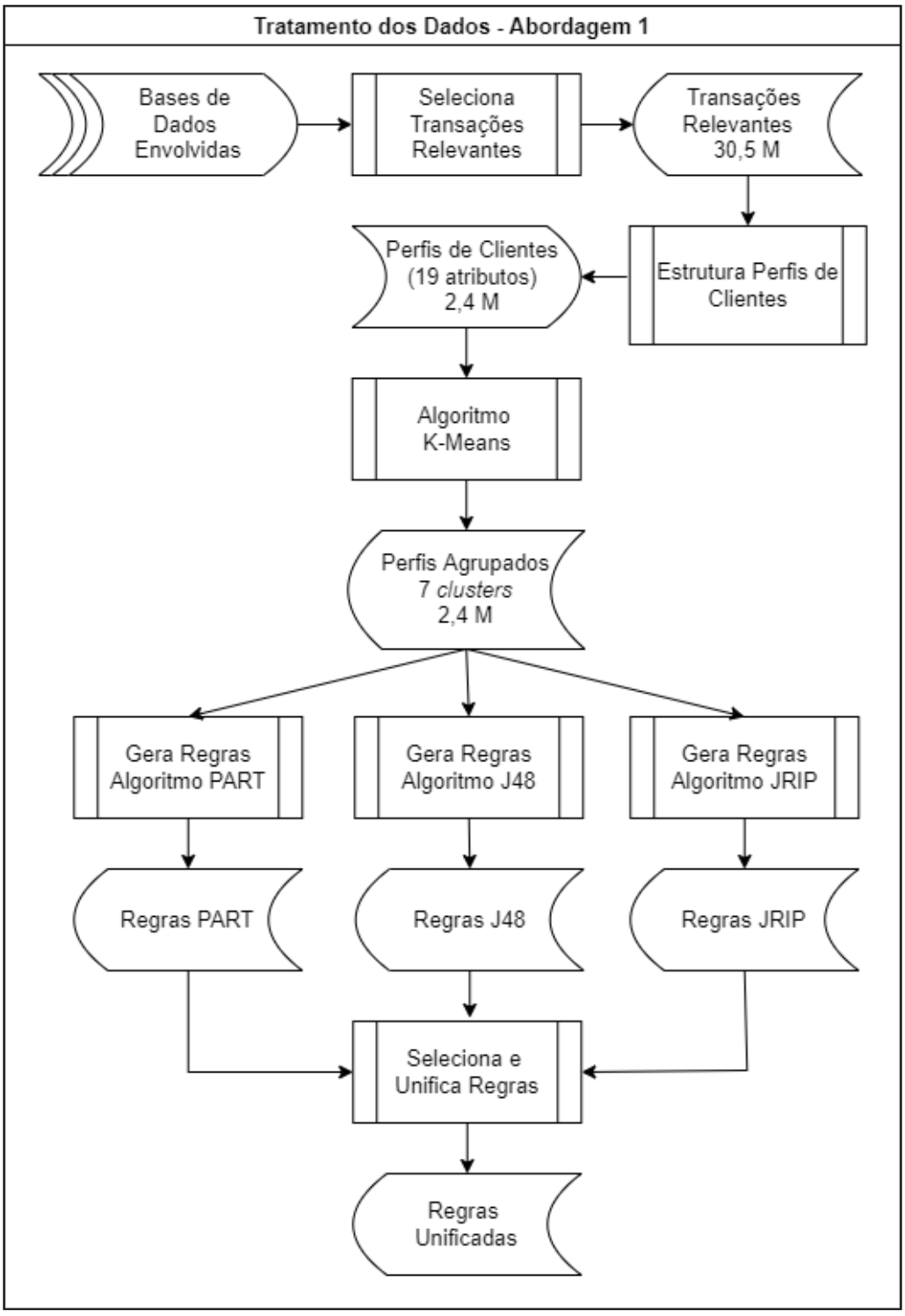
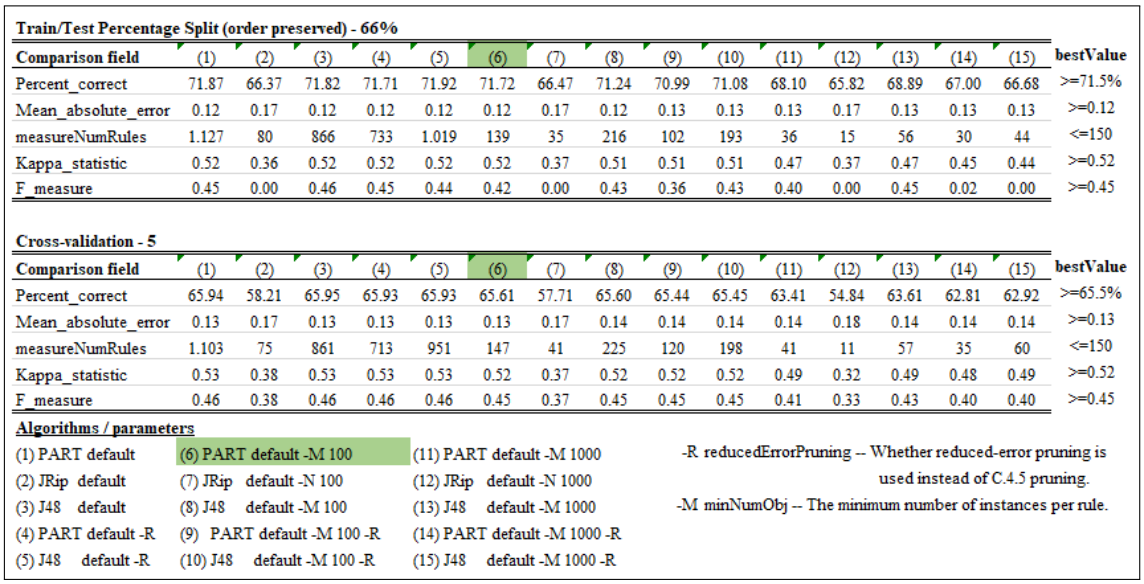


Figura 4.8: Fluxo do Tratamento dos Dados - Abordagem 1

J48 foram executados 22 vezes cada um, variando o número mínimo de instâncias por regra e de folhas na árvore (-M 2 até -M 40000). A cada par de execuções em uma foi utilizada a opção "*reduceErrorPruning*", buscando compressão da árvore gerada, conforme mostra a Tabela 4.4.

As Figuras 4.11 e 4.12 mostram os resultados obtidos. Era esperado que o percentual



Train/Test Percentage Split (order preserved) - 66%

Comparison field	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	bestValue
Percent_correct	71.87	66.37	71.82	71.71	71.92	71.72	66.47	71.24	70.99	71.08	68.10	65.82	68.89	67.00	66.68	>=71.5%
Mean_absolute_error	0.12	0.17	0.12	0.12	0.12	0.12	0.17	0.12	0.13	0.13	0.13	0.17	0.13	0.13	0.13	>=0.12
measureNumRules	1.127	80	866	733	1.019	139	35	216	102	193	36	15	56	30	44	<=150
Kappa_statistic	0.52	0.36	0.52	0.52	0.52	0.52	0.37	0.51	0.51	0.51	0.47	0.37	0.47	0.45	0.44	>=0.52
F_measure	0.45	0.00	0.46	0.45	0.44	0.42	0.00	0.43	0.36	0.43	0.40	0.00	0.45	0.02	0.00	>=0.45

Cross-validation - 5

Comparison field	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	bestValue
Percent_correct	65.94	58.21	65.95	65.93	65.93	65.61	57.71	65.60	65.44	65.45	63.41	54.84	63.61	62.81	62.92	>=65.5%
Mean_absolute_error	0.13	0.17	0.13	0.13	0.13	0.13	0.17	0.14	0.14	0.14	0.14	0.18	0.14	0.14	0.14	>=0.13
measureNumRules	1.103	75	861	713	951	147	41	225	120	198	41	11	57	35	60	<=150
Kappa_statistic	0.53	0.38	0.53	0.53	0.53	0.52	0.37	0.52	0.52	0.52	0.49	0.32	0.49	0.48	0.49	>=0.52
F_measure	0.46	0.38	0.46	0.46	0.46	0.45	0.37	0.45	0.45	0.45	0.41	0.33	0.43	0.40	0.40	>=0.45

Algorithms / parameters

(1) PART default
(2) JRip default
(3) J48 default
(4) PART default -R
(5) J48 default -R
(6) PART default -M 100
(7) JRip default -N 100
(8) J48 default -M 100
(9) PART default -M 100 -R
(10) J48 default -M 100 -R
(11) PART default -M 1000
(12) JRip default -N 1000
(13) J48 default -M 1000
(14) PART default -M 1000 -R
(15) J48 default -M 1000 -R

-R reducedErrorPruning -- Whether reduced-error pruning is used instead of C.4.5 pruning.
-M minNumObj -- The minimum number of instances per rule.

Figura 4.9: Experimento de Geração de Regras (atributos numéricos)

Tabela 4.3: Exemplo da Discretização Realizada

Atributo 10 - Quantidade de Movimentos Realizados no Período	
Faixa	**% do Total do atributo**
<= 3	31%
>= 4, <= 5	32%
>= 6	37%
Atributo 12 - Quantidade de Movimentos na faixa de valor (500,00 - 1.500,00]	
Faixa	**% do Total do atributo**
= 0	31%
>= 1, <= 3	36%
>= 4	32%

Tabela 4.4: Parâmetros Utilizados no Experimento

Num	Parâmetros	Num	Parâmetros	Num	Parâmetros	Num	Parâmetros
1	-M 2	7	-M 3000	13	-M 10000	19	-M 30000
2	-M 2 -R	8	-M 3000 -R	14	-M 10000 -R	20	-M 30000 -R
3	-M 100	9	-M 5000	15	-M 15000	21	-M 40000
4	-M 100 -R	10	-M 5000 -R	16	-M 15000 -R	22	-M 40000 -R
5	-M 1000	11	-M 7000	17	-M 20000		
6	-M 1000 -R	12	-M 7000 -R	18	-M 20000 -R		

de acerto nos testes realizados fosse inversamente proporcional ao aumento da quantidade mínima de instâncias para cobertura das regras ou para as folhas da árvore. Quanto

Train/Test Percentage Split (order preserved) - 66%

Comparison field	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	bestValue
Percent_correct	73.29	70.87	73.28	73.26	73.32	73.26	70.94	73.25	73.22	73.25	72.97	70.63	72.92	72.76	73.00	>=73%
Mean_absolute_error	0.12	0.14	0.12	0.12	0.12	0.12	0.14	0.12	0.12	0.12	0.12	0.14	0.12	0.12	0.12	<=0.12
measureNumRules	1.916	80	1.055	1.216	1.311	274	80	290	182	273	85	39	115	62	91	<=150
Kappa_statistic	0.57	0.51	0.57	0.57	0.57	0.57	0.51	0.57	0.57	0.57	0.57	0.51	0.57	0.56	0.57	>=0.57
F_measure	0.46	0.17	0.46	0.46	0.46	0.46	0.15	0.46	0.45	0.46	0.44	0.16	0.44	0.45	0.48	>=0.46

Cross-validation - 10

Comparison field	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	bestValue
Percent_correct	70.35		70.35	70.34	70.34	70.31		70.28	70.28	70.26	70.01	65.24	69.94	69.88	69.80	>=70%
Mean_absolute_error	0.12		0.12	0.12	0.11	0.12		0.12	0.12	0.12	0.12	0.14	0.12	0.12	0.12	>=0.12
measureNumRules	2.078		2.223	1.398	2.913	290		669	229	623	103	29	249	75	203	<=150
Kappa_statistic	0.57		0.57	0.57	0.57	0.57		0.57	0.57	0.57	0.57	0.48	0.57	0.57	0.57	>=0.57
F_measure	0.69		0.69	0.57	0.69	0.69		0.69	0.69	0.69	0.69	0.62	0.69	0.69	0.68	>=0.69

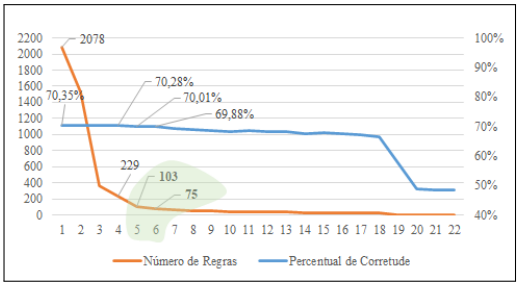
Algorithms / parameters

(1) PART default
(2) JRip default
(3) J48 default
(4) PART default -R
(5) J48 default -R
(6) PART default -M 100
(7) JRip default -N 100
(8) J48 default -M 100
(9) PART default -M 100 -R
(10) J48 default -M 100 -R
(11) PART default -M 1000
(12) JRip default -N 1000
(13) J48 default -M 1000
(14) PART default -M 1000 -R
(15) J48 default -M 1000 -R

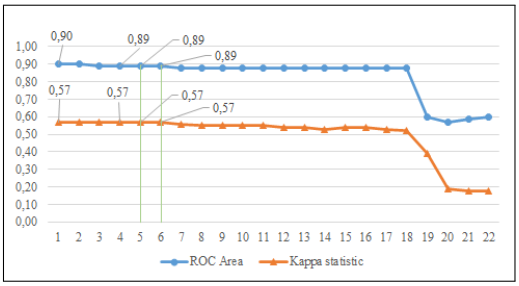
-R reducedErrorPruning -- Whether reduced-error pruning is used instead of C.4.5 pruning.
-M minNumObj -- The minimum number of instances per rule.

Figura 4.10: Experimento de Geração de Regras (atributos nominais)

mais regras geradas maior o nível de acerto. No entanto, a queda acentuada no início da curva, com a criação de um cotovelo, mostra opções de escolha de um menor número de regras sem perda acentuada de qualidade, fato reforçado pela estabilidade das curvas das métricas Receiver Operating Characteristic (ROC) e Kappa.

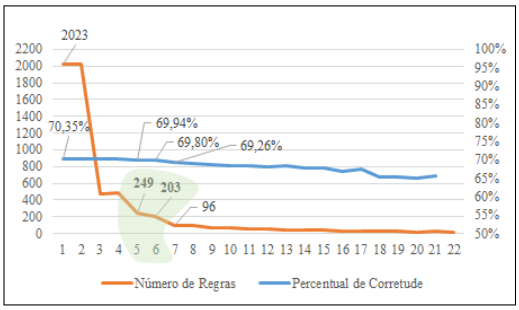


(a) Número de regras e percentual de corretude

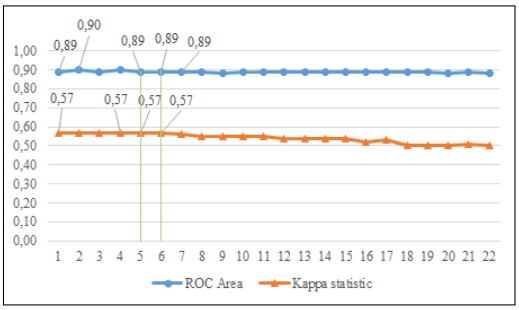


(b) Área ROC e estatística Kappa

Figura 4.11: Algoritmo PART



(a) Número de regras e percentual de corretude



(b) Área ROC e estatística Kappa

Figura 4.12: Algoritmo J48

Diante dos resultados apresentados com este experimento e, principalmente, do apren-

dizado do processo, é possível definir um procedimento automático a ser seguido para geração dos perfis de clientes e das sugestões de regras a serem utilizadas pelos agentes do sistema. Os pontos em destaque na Figuras 4.11(a) e 4.12(a) podem ser utilizados para geração das regras. Uma análise nas regras geradas pelos algoritmos J48 e PART, utilizando os parâmetros do melhor resultado, mostrou uma melhor qualidade considerando completude e ausência de redundância [2].

4.3.3 Algoritmo de Geração dos *Clusters* e Regras - Abordagem 2

A análise da base de dados identificou que menos de 10% dos clientes são pessoas jurídicas (empresas comerciais, industrias, governos etc.), no entanto são responsáveis por mais de 90% dos valores totais envolvidos. Desta forma, visando uma melhor especialização dos *clusters*, ficou evidente que o mais adequado seria realizar a divisão da base de dados em duas: uma constituída somente por clientes singulares e outra com outro tipo de clientes. Com esta divisão da base de dados, todo o procedimento descrito nas secções anteriores e resumido na Figura 4.8 seria executado separadamente para cada base, evidenciando um nível de esforço envolvido na tarefa podendo potencializar o risco de falha no processo. Desta forma, o processo de geração dos *clusters* resultou na formulação e implementação no Algoritmo 1 [3, 4, 5, 6].

Para esta nova estratégia, de cada base foram excluídas as transações cujas características a tornam imunes à prática de branqueamento de capitais, tais como: tarifas, comissões, juros, impostos etc. Na maioria dos casos transações geradas pelos próprios sistemas de negócio do banco. A base final resultou em 30,5 milhões de transações relevantes.

O período medido está diretamente relacionado à natureza do negócio envolvido, apresentando a duração máxima possível, por exemplo trimestral, semestral, anual etc. Neste caso um ano de transações foi utilizado para a geração do perfil de comportamento transacional. Os atributos efetivamente utilizados na aprendizagem visando a formação dos perfis estão listados a seguir e são tratados em Algoritmo 1 - linha 3:

1. Tipo de Pessoa;
2. Idade da Conta (em anos);
3. Quantidade de Serviços Utilizados no Período;
4. Quantidade de Movimentos Realizados no Período;
5. Quantidade de Movimentos na faixa de valor [0,00 – 500,00];
6. Quantidade de Movimentos na faixa de valor (500,00 – 1.500,00];
7. Quantidade de Movimentos na faixa de valor (1.500,00 – 5.000,00];
8. Quantidade de Movimentos na faixa de valor (5.000,00 – 10.000,00];
9. Quantidade de Movimentos na faixa de valor (10.000,00 – 90.000,00];
10. Quantidade de Movimentos na faixa de valor (90.000,00 – inf];
11. Percentual do Montante de Débitos sobre o Montante de Créditos;
12. Percentual da Quantidade de TEDs sobre a Quantidade de Créditos;

13. Percentual da Quantidade de DOCs sobre a Quantidade de Créditos.

Os atributos 5 a 13 mantiveram a discretização descrita na secção anterior e a tabela de perfis dos clientes ativos no ano analisado ficou com 2,4 milhões de linhas, cada linha representando univocamente a tripla cliente, agência e conta (Algorithm 1 – linha 3). Esta tabela foi utilizada num processo de aprendizado indutivo não-supervisionado com *clustering*, para formação de grupos de clientes com características semelhantes e mutuamente exclusivos.

Assim, o passo seguinte é encontrar o melhor número de *clusters k* que represente os diferentes tipos de clientes. Para fazer isso, todos os *k* de 2 a 11 (número de atributos menos 1 – linha 4) foram executados. Para cada *k* (Algorithm 1 – linha 5), foi executada a seguinte sequência: encontrar *k clusters* (usando K-means, representado como *algc*), executar dois classificadores — PART (*algr1*) e J48 (*algr2*) —, assumindo a identificação dos *clusters* como as classes, memorizando os erros encontrados (Algoritmo 1 – linhas 6-10)

Após executar o processo para todos os k's, a função EncontraMenorErro pesquisa os vetores e seleciona aquele que corresponde ao erro mínimo (Algoritmo 1 – linha 13), sendo o resultado final do processo os *Clusters* e Regras correspondentes (Algoritmo 1 – linha 14). O modelo de *cluster* utilizado mostrou-se adequado ao permitir a interpretação dos agrupamentos sem a utilização de sofisticados esquemas de visualização ([31]).

Algorithm 1 Processo de Aprendizagem

```
 1: Input : TR                                          ▷ TR=Transações Relevantes
 2: procedure PROCESSOAPRENDIZAGEM( )
 3:     PerfisDeClientes ← Seleção de Atributos de TR
 4:     k ← Número de Atributos de PerfisDeClientes - 1
 5:     while k > 1 do
 6:         Clusters(k) ← Clustering (algc (PerfisDeClientes, k))
 7:         Regras1(k) ← CriaRegras (algr1 (Clusters (k)))
 8:         Erro1(k) ← CalculaErroClassificação (Regras1(k))
 9:         Regras2(k) ← CriaRegras (algr2 (Clusters (k)))
10:         Erro2(k) ← CalculaErroClassificação (Regras2(k))
11:         k ← k - 1
12:     end while
13:     Idx ← EncontraMenorErro (Erro1, Erro2)
14:     returns Clusters(Idx), Regras1(Idx), Regras2(Idx)
15: end procedure
```

Os resultados obtidos com as execuções do algoritmo confirmaram a expectativa, apresentando *clusters* com as características do estudo anterior (abordagem 1), agora especializado por tipo de pessoa, qual seja: cinco *clusters* para o tipo de pessoa singular e quatro *clusters* para os demais tipos de cliente.

A Figura 4.13 resume esta segunda abordagem descrita acima com a separação das bases de dados e a organização do processo com a definição do Algoritmo 1. O processo assinalado no fluxo como Definição do Nível de Risco está descrito na secção seguinte.

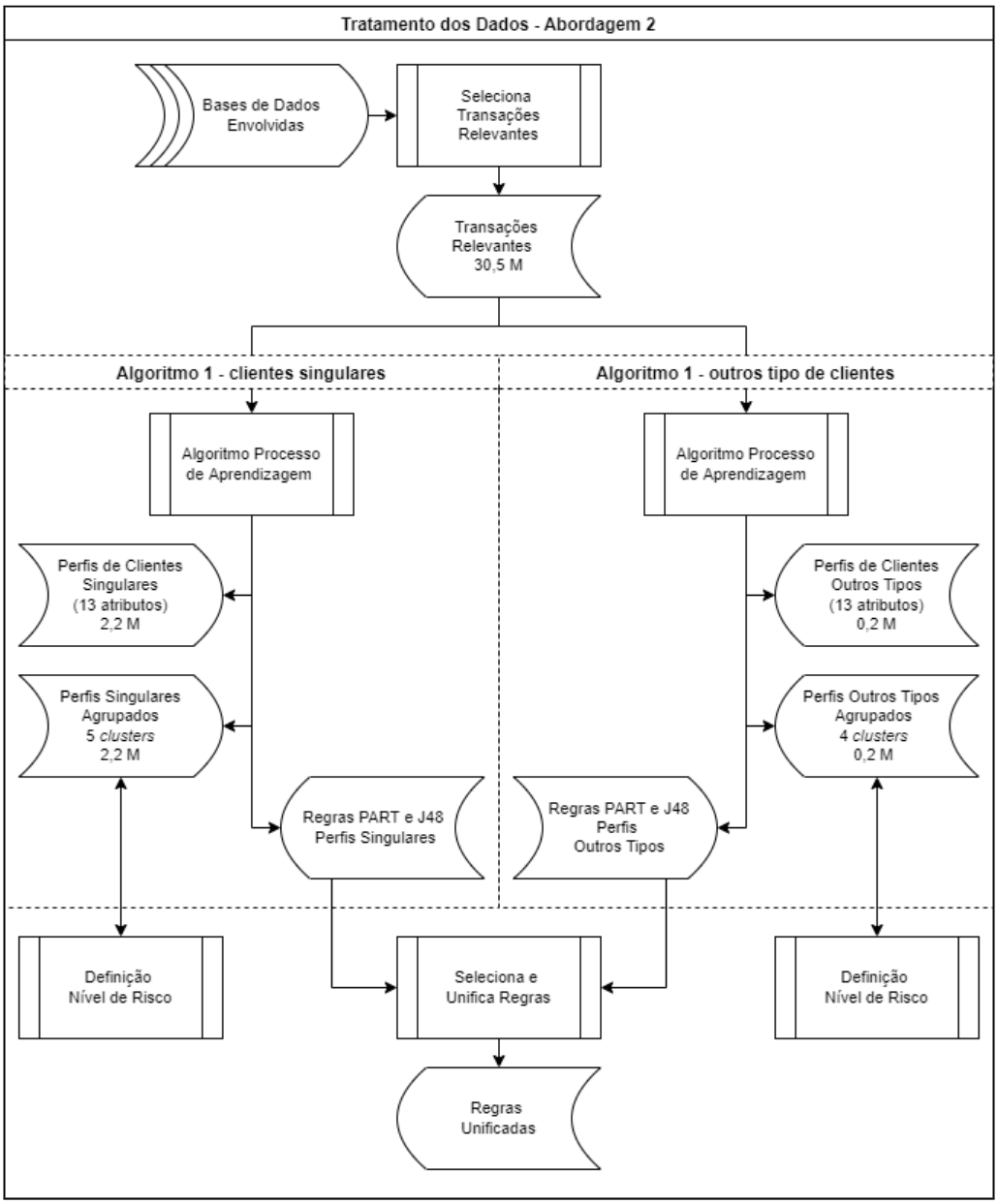


Figura 4.13: Fluxo do Tratamento dos Dados - Abordagem 2

Conforme já comentado o nível de acerto na avaliação das regras geradas, ficou em torno de 99% para ambos os segmentos de clientes, no entanto, foi possível observar no conjunto de original de regras geradas algumas que, apesar de corretas, isoladamente não aplicáveis ao tema Branqueamento de Capitais (BC). Como exemplo é possível citar regras cujos atributos envolvidos se referiam apenas a quantidade de serviços utilizados e idade da conta. Estes elementos não são suficiente para indicar um perfil como suspeito. Foi possível perceber, também, que na quase totalidade desses casos os citados atributos integravam outras regras, tornando-as mais completas e coerentes. A etapa Seleciona e Unifica Regras, mostrada na Figura 4.13 ressalta a necessidade e importância de efetuar a análise de todas as regras geradas.

4.4 Abordagem Conservadora em Relação ao Nível de Risco

A análise dos *clusters* gerados, para os dois segmentos de clientes, permitiu a identificação de características que possibilitam classificá-los conforme o risco da prática de BC. Desta forma, um grupo de clientes com grande movimentação de valores, com transferência integral para outras instituições financeiras, pode ser classificado como de alto risco; já um grupo com movimentação constante de valores próximos do limite de comunicação aos órgãos reguladores pode ser classificado como de risco moderado. Desta análise resultou a classificação mostrada na Tabela 4.5.

Tabela 4.5: Classificação dos Perfis Gerados

Perfil	**Pessoa Singular**	**Outro Tipo de Cliente**
1. Baixa Utilização	ClusterS4	ClusterO4
2. Cliente Padrão	ClusterS2	ClusterO1
3. Risco 1 (baixo)	ClusterS1	-
4. Risco 2 (médio)	ClusterS5	ClusterO3
5. Risco 3 (alto)	ClusterS3	ClusterO2

Com esta classificação é possível definir uma melhor estratégia, oferecendo tratamento diferenciado aos grupos de clientes, conforme seu nível de risco. Apesar do ótimo nível de acerto obtido na avaliação das regras geradas, em torno de 99% para ambos os segmentos de clientes, um por cento de erro representa mais de 24 mil transações e não podem ser desprezadas. A matriz de confusão gerada pelos algoritmos identificou as regras que por serem aplicáveis a dois ou mais grupos de clientes compõem o 1% de erro mencionado. Ou seja, regras classificam clientes como pertencentes a mais de um perfil. Com o propósito de mitigar esta falha, a opção foi reclassificar os perfis que não representam risco ou têm baixo risco (perfis 1, 2 e 3 de pessoa singular e perfis 1 e 2 de outro tipo de cliente), conforme mostra a Tabela 4.6.

Dessa forma, a transação cujo cliente pertencer a um desses 3 grupos será reclassificada, somente para efeito de análise, sendo submetidas as regras do perfil de risco mais elevado que satisfaçam a condição. Por exemplo, na base de dados utilizada para *data mining*, 33 regras classificam os clientes pessoas singulares como pertencentes aos perfis Risco 2 e Risco 3, correspondendo a 1,85% do total. Contudo, estas regras também classificam 0,06% de clientes originalmente pertencentes ao perfil padrão. A reclassificação consiste em, durante o processo de busca por transação suspeita, considerar esses clientes do perfil padrão como pertencentes aos grupos de risco, sem modificar a classificação original.

O método principal adotado para a busca por transações suspeitas consiste em re-

troagir sempre um mês a partir da data da análise, desta forma será sempre utilizado o comportamento do cliente durante um mês de transações para efeito de comparação com os perfis históricos. Para os perfis de risco, serão sempre utilizados para os cálculos o maior valor mensal encontrado para cada atributo e, para os demais perfis, será utilizado como limite o valor total anual, conforme consta da Tabela 4.6. Por exemplo, um cliente padrão, perfil 2, usará sempre como valor referência sua movimentação anual, ou seja, ele será considerado suspeito se em um mês ele ultrapassar a movimentação histórica de um ano. Porém, um cliente classificado como pertencente ao grupo de risco (perfis 3 a 5 de pessoa singular e perfis 4 e 5 de outro tipo de cliente) têm como valor referência a movimentação histórica mensal, isto é, basta que a movimentação do último mês ultrapasse o referencial histórico para ser considerado suspeito.

Tabela 4.6: Perfis Gerados - Informações de Controle

Perfil	**Pessoa Singular**	**Outro Tipo de Cliente**	**Reclassifica Perfil**	**Valor Limite**
1. Baixa Utilização	ClusterS4	ClusterO4	Sim	Total anual
2. Cliente Padrão	ClusterS2	ClusterO1	Sim	Total anual
3. Risco 1 (baixo)	ClusterS1	-	Sim	Máximo mensal
4. Risco 2 (médio)	ClusterS5	ClusterO3	Não	Máximo mensal
5. Risco 3 (alto)	ClusterS3	ClusterO2	Não	Máximo mensal

A base de conhecimento criada, composta por regras geradas no processo de *data mining*, regras criadas com base em leis e diretrizes gerais, e por regras formadas com base no conhecimento dos Analistas de ABC deve ser a mais estável possível, considerando que se transformará em crenças a serem utilizadas pelos agentes, possibilitando a tomada de decisão consistentes e coerentes. No entanto, considerando o apetite e tolerância ao risco[6] praticados pela instituição financeira utilizadora desta abordagem, ela pode desejar, para uma determinada análise, ser mais ou menos rigorosa quanto aos valores limites. Esta prática levaria, inevitavelmente, a modificações nas regras, quebrando a desejada estabilidade.

Para contornar esta situação, foi parametrizado um valor que representa um percentual intitulado Margem Adicional de Risco (MAR), cuja atribuição ficará a critério da instituição financeira e que, aplicado sobre os valores limites, reduzem-no ou ampliam-no pelo percentual indicado [4, 5, 6].

A Margem Adicional de Risco (MAR) é definida como:

$$mar_{i,C} = \begin{cases} V_{i,C}, & P(C) = 4 \vee 5 \\ \frac{MAR}{100} \times V_{i,C}, & P(C) = 1 \vee 2 \vee 3 \end{cases} \tag{4.2}$$

[6]O Apetite ao Risco é o nível de risco que uma organização está disposta a aceitar enquanto persegue seus objetivos. Normas brasileiras como ISO 27001, ISO 27005 e ISO 31000, abordam o assunto.

onde: $mar_{a,C}$ é a margem adicional de risco considerada para o atributo *a* do cliente *C*; *MAR* é o valor da margem de risco a ser aplicada para os atributos; $V_{i,C}$ é o valor referência dos atributos *i* do cliente *C*; *P(C)* obtém o perfil do cliente *C*.

Uma Margem Adicional de Risco (MAR) maior do que zero provoca a execução de uma análise mais conservadora e rigorosa, devendo produzir uma elevação na quantidade de perfis suspeitos. Valores da MAR acima de 100 provocam uma análise mais liberal, tendo como consequência esperada uma redução na quantidade de perfis suspeitos. Por outro lado, uma MAR menor que 100 executa uma análise mais conservadora, resultando numa quantidade maior de perfis suspeitos. A MAR igual a 100 não modifica os valores apurados e geram a quantidade real de casos suspeitos. Os clientes com perfis de risco médio e alto não são afetados pela MAR, gerando sempre a quantidade original de clientes suspeitos. A Equação 4.2 formaliza a Margem Adicional de Risco (MAR) e a Tabela 4.7 resume os efeitos da aplicação da MAR.

Tabela 4.7: Atuação da MAR sobre o Resultado da Análise

MAR	Tipo de Análise	Efeito nos Perfis		Resultado Obtido
		Baixa Utilização, Cliente Padrão e Baixo Risco	Risco Médio e Risco Alto	
> 100	Liberal	Limite Ampliado	Valor Limite	Menos suspeitos
$= 100$	Padrão	Valor Limite	Valor Limite	Quantidade Real
< 100	Conservadora	Limite Reduzido	Valor Limite	Mais suspeitos

A Tabela 4.8 mostra como a MAR é aplicada a um determinado perfil. Caso o perfil em análise esteja classificado como sendo 1 ou 2, conforme mostrado na Tabela 4.6, a quantidade máxima anual apurada para o analisado será utilizada como referência de comparação. Este valor referencial será modificado ou não, conforme o percentual atribuído a MAR, podendo ser menor, igual ou superior a 100%. A MAR também atua sobre o perfil 3 (Risco baixo), porém, o referencial será a quantidade máxima mensal apurada para o analisado. Importante ressaltar que a MAR não será utilizada nos perfis 4 e 5 que são os de de maior risco.

A título de exemplo, admitindo a análise do atributo 4 (Quantidade de Movimentos Realizados no Período) e os valores mostrados na Tabela 4.8, é possível perceber que uma MAR de 100% significa a aplicação de uma análise padrão, ou seja, as quantidades máximas apuradas serão utilizadas na decisão de suspeição. Os perfis 1 e 2 são comparados com o limite anual de 450 transações e os demais perfis serão confrontados com o limite mensal de 50 transações. No caso de uma MAR de 110% ela será aplicada sobre o limite anual para avaliação dos perfis 1 e 2, que serão considerados suspeitos se a movimentação em análise for superior a 495 transações. Com a mesma MAR o perfil 3 (risco baixo) será considerado suspeito somente a partir de 55 transações. Com uma MAR de 90% (conservadora), a aplicação é a mesma, sendo que diminui os limites mensais ou anuais,

Tabela 4.8: Aplicação da MAR em um Atributo

Cliente X	**Mensal**	**Anual**
Quantidade Máxima do Atributo	50	450

	Análise Liberal		**Análise Padrão**		**Análise Conservadora**	
Perfil do Cliente X	**MAR=110%**	**Suspeito se Atributo**	**MAR=100%**	**Suspeito se Atributo**	**MAR=90%**	**Suspeito se Atributo**
Baixa Utilização	450+45	>495	450+0	>450	450-45	>405
Cliente Padrão	450+45	>495	450+0	>450	450-45	>405
Risco 1 (baixo)	50+5	>55	50+0	>50	50-5	>45
Risco 2 (médio)	50	>50	50	>50	50	>50
Risco 3 (alto)	50	>50	50	>50	50	>50

conforme o caso. Ou seja, os perfis 1 e 2 passam admitir o limite de 405 transações e o perfil 3 passa admitir o limite de 45 transações. Qualquer que seja a MAR utilizada os perfis 4 e 5, classificados como de risco médio e alto, sempre serão avaliado numa comparação com o valor mensal calculado de 50 transações.

4.5 Índice de Suspeição

Por mais agilidade que a adoção de novas tecnologias possa trazer para o processo de Anti-Branqueamento de Capitais (ABC) um estoque de perfis a serem analisados sempre será uma variável que, em algum momento, pode impor ao processo a necessidade de priorização na análise. Um acumulo de casos complexos de perfis suspeitos que necessitem da intervenção do Analista de BC pode ocorrer por vários motivos: regras que necessitam de revisão; crescimento natural da volumetria de transações; redução no número de analistas; etc.

Numa situação desta uma pergunta que sempre é feita: diante do estoque, quem deve ser analisado primeiro. Com o objetivo de apresentar a relação de perfis suspeitos com uma ordenação em função do risco, foi estabelecida a possibilidade de atribuição de peso para cada um dos onze atributos utilizados para formação dos *clusters*. Este peso é um inteiro no intervalo de um a nove que será multiplicado pela ponderação dos valores dos atributos de perfil em análise. A soma dos valores dos atributos define o índice de suspeição do perfil. A ordenação decrescente destes índices fornece uma relação de prioridade dos perfis que, em última análise, é uma ordenação por um critério de risco.

Neste trabalho, três dos atributos foram considerados os mais importantes e receberam peso 2. Esses atributos são: a Quantidade de Movimentos na faixa de valores (5.000,00 - 10.000,00], o Percentual de Transferências Eletrónicas Disponíveis (TED) e o Percentual de Documento de Ordem de Crédito (DOC). O primeiro, por ser um intervalo cujo valor máximo é explicitamente estabelecido pelo Banco Central do Brasil (BACEN) como passível de comunicação àquele órgão, sempre que este valor atingido numa transação.

Razão pela qual os fraudadores dividem as operações em valores menores e os outros dois porque indicam transferências monetárias para outras instituições financeiras. Estes atributos contribuem fortemente para a formação de perfis de risco. A Equação 4.3 formaliza o cálculo do Índice de Suspeição (IS) e a Tabela 4.9 mostra os atributos e seus respetivos pesos [4, 5, 6].

$$IS(C) = \sum_{k=1}^{na} \frac{v_k(C)}{mx_k(C)} \times p_k \quad , mx_k > 0 \wedge p_k > 0 \tag{4.3}$$

onde: $IS(C)$ é o índice de suspeição do cliente C; na é o número de atributos que caracteriza um perfil; v_k é o valor do atributo k; mx_k é o maior valor mensal ocorrido no período de formação do perfil k; p_k é o peso atribuído ao atributo k.

Tabela 4.9: Pesos dos Atributos Utilizados no Cálculo do Índice de Suspeição

Atributo	**Peso**
1. Quantidade de Serviços Utilizados no Período	1
2. Quantidade de Movimentos Realizados no Período	1
3. Quantidade de Movimentos na faixa de valor [0,00 – 500,00]	1
4. Quantidade de Movimentos na faixa de valor (500,00 – 1.500,00]	1
5. Quantidade de Movimentos na faixa de valor (1.500,00 – 5.000,00]	1
6. Quantidade de Movimentos na faixa de valor (5.000,00 – 10.000,00]	2
7. Quantidade de Movimentos na faixa de valor (10.000,00 – 90.000,00]	1
8. Quantidade de Movimentos na faixa de valor (90.000,00 – inf]	1
9. Volatilidade de Débito	1
10. Percentual de Transferência Eletrônica Disponível (TED)s	2
11. Percentual de Documento de Ordem de Crédito (DOC)s	2

A Tabela 4.10 apresenta exemplos de cálculo do IS em três cenários:

1. Utilização Máxima - neste cenário os limites armazenados e as quantidades em análise reproduzem um perfil de alta utilização dos serviços disponíveis, transações realizadas e de grande movimentação de recursos financeiros. Comprovado pelas quantidades indicadas nas colunas *qtdServ* (quantidade de serviços), *qtdMov* (quantidade de movimentações) e nas duas maiores faixas de valores. O período em análise mostra uma variação maior que o comportamento conhecido;
2. Utilização Intermediária - o perfil em análise neste cenário apresenta comportamento bem próximo do padrão conhecido, quando comparadas as quantidades em análise e o comportamento conhecido. Seria selecionado como suspeito pelo fato de estar acima dos limites conhecidos;
3. Situação Atípica - o cenário procura representar um caso bastante comum, no qual a movimentação, repentinamente, muda de padrão, principalmente nos atributos que

sinalizam comportamento de risco. Os atributos *fxVlr4* (faixa de valor 4, de 5.000 a 10.000), *pctTED* e *pctDOC* (percentuais de tipos de transferências eletrónicas) são exatamente os atributos que recebem pesos na simulação apresentada.

Objetivando demonstrar o efeito da aplicação dos pesos no resultado final do IS a Tabela 4.10 apresenta um cálculo parcial do índice sem aplicar os pesos e em seguida os cálculo final com os pesos, destacando os atributos envolvidos. É possível perceber que o IS final apresenta pouca alteração entre o cálculo intermediário e o resultado final para o cenário intermediário. No cenário de utilização máxima a variação é maior, indicando um desvio merecedor de destaque. No entanto, o IS final pontua fortemente o perfil do cenário atípico, quando comparado com os demais cenários, conforme esperado. Desta forma, ordenando o resultado a prioridade de análise mais aprofundada por parte de um humano seria do perfil do cenário atípico, seguido pelo de maior utilização e, por último, o de utilização intermediária.

Tabela 4.10: Exemplo do Cálculo do Índice de Suspeição

	Limites do Perfil em Análise										
Cenários	**qtdServ**	**qtdMov**	**fxVlr1**	**fxVlr2**	**fxVlr3**	**fxVlr4**	**fxVlr5**	**fxVlr6**	**volatDeb**	**pctTED**	**pctDOC**
Máximo	23	7567	5737	578	709	183	299	61	180,11	120,15	39,86
Intermediário	17	1378	680	266	245	69	60	58	190,23	130,42	26,85
Atípico	8	49	23	8	8	9	1	0	102,15	1,5	0
	Quantidades do Mês para o Perfil em Análise										
Cenários	**qtdServ**	**qtdMov**	**fxVlr1**	**fxVlr2**	**fxVlr3**	**fxVlr4**	**fxVlr5**	**fxVlr6**	**volatDeb**	**pctTED**	**pctDOC**
Máximo	26	9620	7795	824	525	192	202	82	128,93	184,52	42,35
Intermediário	16	1663	702	356	432	72	71	30	160,87	120,87	18,32
Atípico	8	92	25	12	11	37	6	1	105,31	98,42	9,03
	Cálculo do IS sem Pesos											
Cenários	**qtdServ**	**qtdMov**	**fxVlr1**	**fxVlr2**	**fxVlr3**	**fxVlr4**	**fxVlr5**	**fxVlr6**	**volatDeb**	**pctTED**	**pctDOC**	**IS**
Máximo	1,13	1,27	1,36	1,43	0,74	1,05	0,68	1,34	0,72	1,54	42,35	53,60
Intermediário	0,94	1,21	1,03	1,34	1,76	1,04	1,18	0,52	0,85	0,93	0,68	11,48
Atípico	1,00	1,88	1,09	1,50	1,38	4,11	6,00	1,00	1,03	65,61	9,03	93,62
	Cálculo do IS com Pesos											
Cenários	**qtdServ**	**qtdMov**	**fxVlr1**	**fxVlr2**	**fxVlr3**	**fxVlr4**	**fxVlr5**	**fxVlr6**	**volatDeb**	**pctTED**	**pctDOC**	**IS**
Máximo	1,13	1,27	1,36	1,43	0,74	2,10	0,68	1,34	0,72	3,07	84,70	98,53
Intermediário	0,94	1,21	1,03	1,34	1,76	2,09	1,18	0,52	0,85	1,85	1,36	14,13
Atípico	1,00	1,88	1,09	1,50	1,38	8,22	6,00	1,00	1,03	131,23	18,06	172,38

4.6 Conclusão

Com o tratamento dos realizado, a abordagem de perfis dos clientes definida, os perfis agrupados por nível de risco, processo de geração de regras sistematizado em um algoritmo, técnicas auxiliares com Margem Adicional de Risco e Índice de Suspeição definidas, resta implementar, testar e validar os resultados. Estas etapas estão descritas nos próximos capítulos.

Capítulo 5

Modelagem e Implementação de um Sistema Multi-agentes para ABC

5.1 Contexto

A capacidade de readaptação do *modus operandi* dos fraudadores e a falta de informação sistematizada que permita associar as transações informadas como suspeitas com a comprovação do crime, comentado em Um Possível Risco Sistémico (tópico 2.2.3 - pág. 21), são fatores que dificultam bastante a automatização do processo de prevenção e combate ao crime de Branqueamento de Capitais (BC).

Apesar do presente trabalho não atuar diretamente sobre o risco sistémico citado anteriormente, posto que isto demandaria uma ação sobre todo o sistema financeiro, o resultado, porém, tende a mitigá-lo, na medida em que busca melhorar o processo interno das instituições financeiras, permitindo elevar qualitativamente a identificação de casos suspeitos e, dessa forma, contribuir para a melhora da definição de novos normativos e parâmetros de captura.

O Fluxo Genérico de Combate a Fraudes (FGCF), mostrado na Figura 2.5 (pág. 15), apresenta oportunidades de melhoria. A aplicação de um novo fluxo no atual processo de Anti-Branqueamento de Capitais (ABC) das instituições financeiras, mostrado na Figura 2.7 (pág. 17), confirma os pontos críticos já citados: a captura de transações suspeitas e a tarefa de análise executada pelos Analistas de ABC. Estes são os principais pontos abordados neste capítulo.

5.2 Novo Fluxo Genérico de Combate à Fraude

A quase totalidade das soluções tecnológicas existentes actualmente que visam suportar o processo de ABC, concentram-se na parte do processo referente a captura de transações suspeitas, transferindo para o Analista de ABC a tarefa de comprovação ou não da suspeição. Além disso, boa parte dessas soluções são procedurais e fortemente baseadas

em parametrizações oriundas dos normativos publicados. Um grande esforço de inovação tem ocorrido, contudo, a maioria dos algoritmos propostos concentram-se na busca de identificação de transações suspeitas considerando visões parciais e não testados com grandes volumes de dados.

É possível que esta seja a explicação para o fato de entre 317 profissionais dedicados a operações de ABC e técnicos de conformidade, em instituições financeiras de 48 países, incluindo do Brasil e de Portugal, apenas 58% acreditam que os sistemas existentes em suas organizações são capazes de monitorar adequadamente as transações ocorridas nas diversas linhas de negócios. Além disso, somente 49% destes profissionais afirmaram que os citados sistemas de monitoração são capazes de compartilhar informações entre as diversas plataformas de negócio. Estes dados constam do documento *Global Anti-Money Laundering Survey*, divulgado pela *KPMG International Corporate (KPMG)*, em 2014 [75].

A *Dow Jones Risk & Compliance* e a *Swift Financial Crime Compliance* entrevistaram mais de 500 executivos da área de gestão e riscos ao redor do mundo para compor o documento *Global Anti-Money Laundering Survey Results 2017* [48]. Um dos objetivos da pesquisa era mapear as tendências e preocupações dos executivos com ABC. Entre os principais desafios relacionados estão: reduzir o alto número de falsos positivos (46%) e lidar com tecnologias insuficiente, inadequada ou desatualizada (48%).

A empresa de consultoria AlixPartners em seu *2018 Global Anti-Money Laundering and Sanctions Compliance Survey* [7], entrevistou 372 instituições financeiras em 71 países e identificou uma mudança significativa nos desafios apontados pelos entrevistados, quando comparados com os dados de 2017. Em 2018, o primeiro desafio é aplicar avaliação de risco em ABC (54%, era 50% em 2017), sendo segundo desafio o uso de sistemas eficientes de monitoramento de transações para detetar atividades potencialmente suspeitas (48%, era 50% em 2017).

Estas pesquisas reforçam a perceção de que o processo de ABC tem problemas a serem resolvidos. A experiência tem mostrado que este processo sendo praticado como uma instância do fluxo genérico de combate a fraude é ineficiente, pois ele falha em identificar e reportar as transações suspeitas. Esta ineficiência contribui fortemente para a formação do risco sistémico, pois a redução de casos reais interfere na formulação de novas leis e recomendações de qualidade.

A Figura 5.1 mostra uma nova versão do fluxo geral de combate à fraude, cujo objetivo é mitigar o risco sistémico e buscar solução para os problemas acima citados [4]. A criação de perfis dos atores participantes da atividade, baseados no histórico completo das operações realizadas; a substituição dos atuais parâmetros fixos por regras de produção, aprendidas com estes perfis e nas normas e recomendações existentes; a eliminação da utilização das operações mais recentes; abordagem baseada em risco; e o uso de inteligência em alguns pontos do processo eleva qualitativamente o nível na captura de operações suspeitas e, principalmente, da tomada de decisão pelo especialista [99].

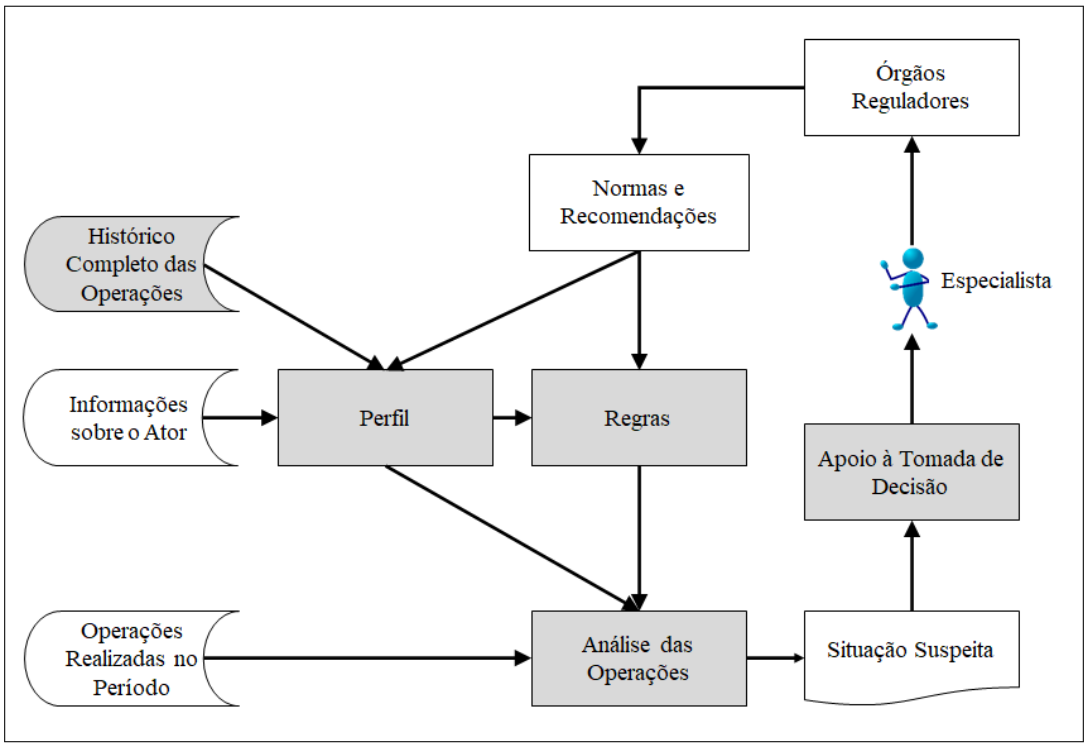


Figura 5.1: Fluxo Genérico de Combate a Fraudes (Proposto)

5.3 Análise e Modelagem de um Sistema Baseado em Agentes

Utilizando o novo fluxo genérico de combate à fraude proposto como uma instância do processo de Anti-Branqueamento de Capitais (ABC), é possível observar que algumas das tarefas a serem automatizadas (pelo menos parcialmente) mantém sintonia com os princípios que definem um Sistemas Multi-Agente (SMA) [128]. É necessário um conjunto de entidades (agentes) com autonomia para realizar tarefas específicas e que mantenham comunicação com as outras entidades, a fim de atingir um objetivo comum. Cada entidade deve ter o seu próprio conhecimento e deve ser capaz de raciocinar e decidir de forma inteligente. Além disso, essas entidades precisam apresentar flexibilidade e escalabilidade [44].

Na tomada de decisão, uma característica torna-se muito importante: a persistência no alcance do objetivo. Caso um agente, por algum motivo, não consiga atingir um objetivo por meio de uma intenção específica, ele deverá reconsiderar o objetivo, considerando o contexto atual. Nesta situação, um agente baseado no modelo BDI (*Belief-Desire-Intention*) é capaz de tentar encontrar um novo curso de ação para atingir o objetivo, descartando-o somente quando é alcançado ou considerado irrelevante [11].

Considerando outras características como: quantidade e relevância dos dados históricos disponíveis que serão utilizados para a tomada de decisão; a necessidade mencionada de, durante o processo, rever e expandir o conhecimento adquirido; a possibilidade de de-

senvolver novos subobjetivos / objetivos; considera-se que BDI seja o modelo adequado para ser aplicado ao caso.

Diante do exposto, o objetivo geral da solução proposta por este estudo pode ser assim descrito:

> *Desenvolver um sistema para dar suporte ao processo de Anti-Branqueamento de Capitais (ABC) em uma instituição financeira. O sistema irá manter atualizado um perfil para cada cliente, baseado em seu histórico de transações, que será utilizado, juntamente com as regras extraídas dos normativos oficiais de combate ao Branqueamento de Capitais (BC), para a captura e sinalização de transações suspeitas processadas pelos diversos sistemas de negócio. O sistema decidirá sobre alguns casos sinalizados e aprenderá com o auxílio que prestará ao Analista de ABC na tomada de decisão dos casos mais complexos. Novas regras para captura de transações suspeitas e as possíveis mudanças em perfis, serão sugeridas.*

A Figura 5.2 mostra a incorporação dos objetivos acima descritos ao fluxo genérico apresentado na Figura 5.1. Os principais aspetos e inovações deste modelo são [6]:

1. O mapeamento do comportamento transacional de cada cliente, considerando atributos específicos que ajudam a definir níveis de risco de lavagem de dinheiro, permitindo a criação de grupos de clientes com risco potencial (A);
2. A modelagem dos processos de captura e análise de transações suspeitas com base em técnicas de inteligência artificial utilizando perfis comportamentais mapeados, além de padrões e recomendações existentes (B);
3. Agente inteligente executando tarefas e auxiliando os analistas humanos no processo de tomada de decisão (C);
4. Uso de estratégias baseadas em risco tanto na captação quanto na tomada de decisão, oferecendo flexibilidade à instituição financeira no exercício de seu apetite ao risco (B e C).

5.3.1 Metodologia de Desenvolvimento Utilizada

Objetivando um nível de documentação adequado e, em última instância, uma qualidade melhor na modelagem do sistema proposto, torna-se necessária a utilização de uma metodologia formal de apoio ao processo de modelagem. Existe uma grande variedade de metodologias propostas para suportar uma abordagem de Engenharia de Software Orientada aos Agentes (*AOSE - Agent-Oriented Software Engineering*), algumas delas implementadas em produtos de mercado, ou grandes ambientes de desenvolvimento. Dessa forma, foram estabelecidas as seguintes premissas para escolha da metodologia a ser utilizada: a) de fácil utilização; b) poucos elementos de representação; 3) documentação acessível,

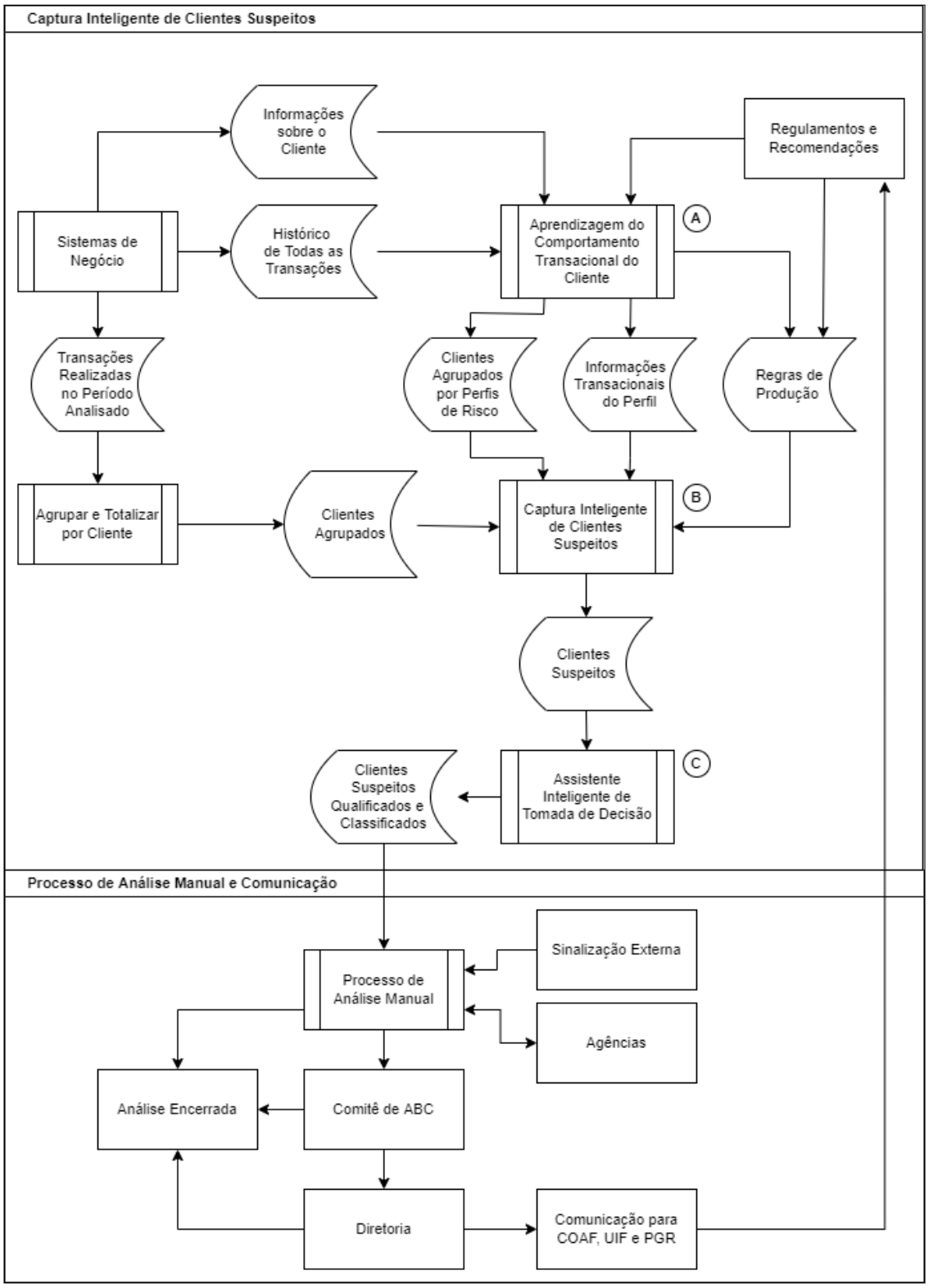


Figura 5.2: Novo Fluxo de ABC

de linguagem direta e, de preferência, com exemplos; 4) ferramenta de apoio de fácil instalação, de preferência para o ambiente Windows.

Wooldridge [128] apresenta em algumas metodologias e entre elas a Prometheus, como sendo simples, com uma rica coleção de modelos e com ferramenta de suporte. Es-

tudo sobre as principais Agent-Oriented Software Engineering (AOSE), apresentado por Winikoff and Padgham [127], Prometheus é citada e boa parte dos exemplos mostrados no trabalho utilizam as notações da metodologia. Bawa and Attri [18] realiza comparação entre cinco metodologias e conclui que os agentes na Prometheus não são simples entidades, eles são compostos por entidades menores com papéis e competências.

A metodologia Prometheus consiste em três fases, mostradas na Figura 5.4:

1. A fase de especificação do sistema, que foca na identificação dos objetivos e funcionalidades básicas do sistema, desenvolvimento de cenários de caso de uso, juntamente com entradas (perceções), dados, atores, papéis e saídas (ações);
2. A fase de projeto da arquitetura, que usa as saídas da fase anterior para determinar quais tipos de agentes serão implementados, desenvolver seus protocolos e como eles irão interagir;
3. A fase de detalhamento do projeto, que refina as capacidades internas de cada agente, especifica os processos e define como serão realizadas suas tarefas dentro do sistema. Nesta fase também são especificados os algoritmos de cada plano de ação, as crenças e detalhamentos dos eventos.

Assim sendo, a metodologia foi testada e comprovou-se ser de fácil utilização, bem documentada[1], principalmente devido ao livro [96], que apresenta bons exemplos e pela ferramenta de apoio chamada *Prometheus Design Tool (PDT)* (um *plugin* do Eclipse). Exemplos práticos de utilização da metodologia foram apresentados por de Silva et al. [41] e numa proposta de solução para decisões táticas na área militar [51].

A Figura 5.3 apresenta um *screenshot* dos elementos gráficos disponíveis na PDT. Sobre as setas faz-se necessária uma explicação adicional pois pode representar subordinação (como no caso dos objetivos e subobjetivos), direção do fluxo de informação (como nas trocas de mensagens, perceções e ações) ou recuperação/geração de dados (como no caso de leitura e/ou gravação de dados).

Padgham and Winikoff [96], autores da metodologia, deixaram claro que a Prometheus é mais aderente a uma plataforma de agentes baseada em BDI, no entanto, somente no momento de geração dos planos isso fica evidente, o que não impede de a metodologia ser utilizada para outras plataformas. A Figura 5.4 apresenta o esquema geral da metodologia, com suas fases e os itens disponíveis na ferramenta PDT. No tópico seguinte deste trabalho serão apresentados os diagramas mais importantes da modelagem da solução proposta e o restante da documentação está no Apêndice B.

5.3.2 Objetivos do Sistema

Para definir os objetivos e subobjetivos do sistema, um dos passos iniciais é analisar a definição feita para o sistema no tópico 5.3 (página 74). Dessa forma, é possível obter a seguinte relação inicial de objetivos (os subobjetivos foram obtidos a partir da pergunta

[1] http://www.rmitagents.com/

[2] As figuras referentes à Metodologia serão mantidas no original em inglês

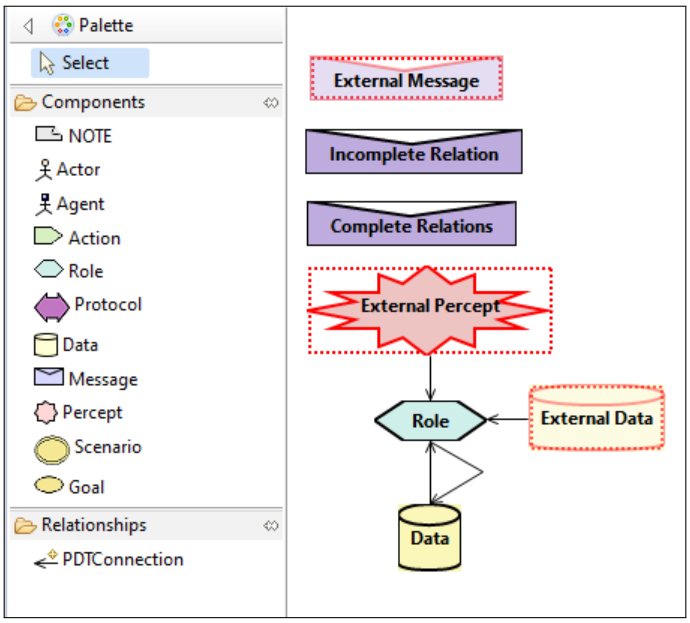
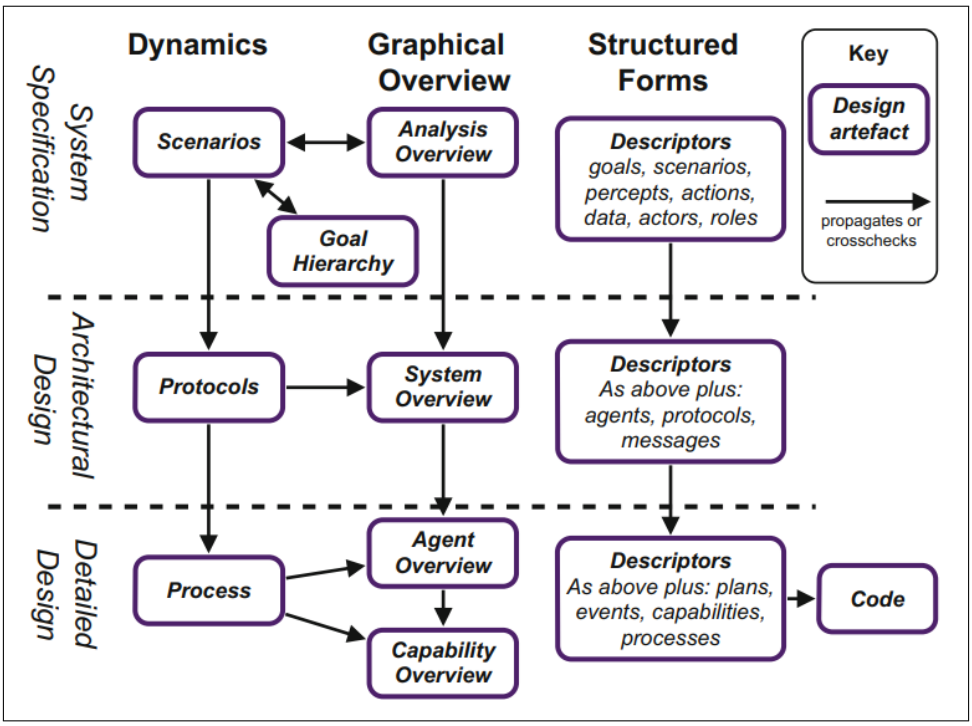


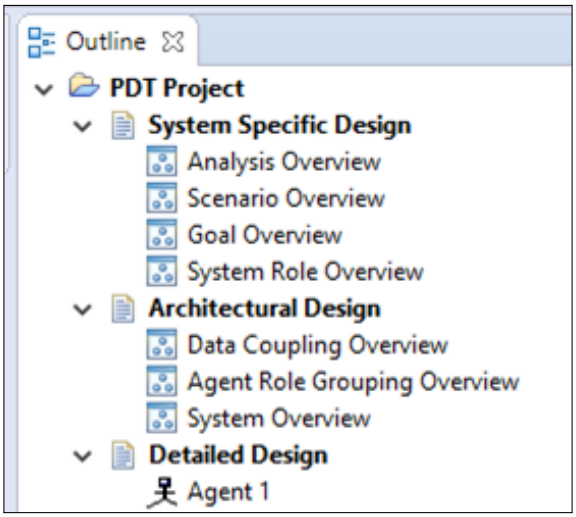
Figura 5.3: PDT: Elementos Gráficos de Representação[2]

"como?" feita para cada objetivo):

1. dar suporte ao processo de anti-branqueamento de capitais;
   1.1. automatizar etapas do processo;
   1.2. dotar de inteligência etapas automatizadas;
   1.3. aprender com a execução do processo;
2. manter um perfil para cada cliente;
   2.1. gerar perfis com base em histórico de transações;
   2.2. confrontar novos perfis com perfis existentes;
   2.3. sugerir novos perfis;
   2.4. efetuar atualizações na base perfis;
3. capturar perfis suspeitos;
   3.1. analisar base histórica de transações;
   3.2. aplicar regras geradas no processo de aprendizagem;
   3.3. aplicar regras extraídas dos normativos oficiais de ABC;
   3.4. aplicar regras extraídas dos perfis dos clientes;
   3.5. guardar transação suspeita;
   3.6. comunicar existência de transação suspeita;
4. sinalizar perfis suspeitos;
   4.1. verificar todos os produtos aplicáveis para cada perfil suspeito;
   4.2. comunicar a existência de perfis suspeitos;
5. decidir sobre casos sinalizados;
   5.1. analisar sinalização à luz do histórico de decisões tomadas;
   5.2. guardar decisão tomada no histórico de decisão;
6. auxiliar na tomada de decisão dos casos mais complexos;
   6.1. decidir sobre os casos sinalizados;



(a) Fases da Metodologia [96]



(b) Implementação na PDT

Figura 5.4: Metodologia Prometheus

6.2. sinalizar casos passíveis de análise do Analista de ABC;

7. aprender com a tomada de decisão nos casos mais complexos;
   7.1. gerenciar a atualização do histórico de decisões tomadas;
   7.2. garantir que exista uma decisão para cada sinalização feita;
   7.3. utilizar histórico de decisões na tomada de decisão;
8. sugerir novas regras incluindo novos perfis;
   8.1. elaborar regras baseadas nos novos perfis encontrados;
   8.2. analisar regras existentes;
   8.3. guardar sugestão de novas regra.

Estes objetivos e subobjetivos foram analisados, ampliados ou agrupados até a obtenção de um modelo adequado. As etapas seguintes vão refinando, ampliando ou fundindo objetivos, gerando novos artefatos, ampliando a definição até que um diagrama geral do sistema torne-se aceitável.

5.3.3 Diagramas do Sistema

Os diagramas de cada uma das etapas estão colocados no Apêndice B, conforme a Tabela 5.1.

Tabela 5.1: Diagramas Gerados pela PDT

Etapas	**Apêndice B**
Objetivos e subobjetivos	Figura B.1
Objetivos agrupados por funcionalidades (*roles*)	Figura B.2
Operação do sistema em cenários	Figura B.3
Interligação das bases de dados	Figura B.4
Criação de agentes por agrupamento de *roles*	Figura B.5
Diagrama geral do sistema	Figura B.6
Detalhamento dos agentes	Figuras B.7 a B.15
Diagrama AUML Gerencia Captura	Figura B.16
Diagrama AUML Atualiza Perfis	Figura B.17
Diagrama AUML Evolui Regras	Figura B.18
Diagrama AUML Valida Sugestões	Figura B.19

A Figura 5.5 mostra uma versão simplificada do diagrama mostrado na Figura B.6 do Apêndice B. Esta versão do diagrama ressalta os agentes, as bases de dados externas (DB), as bases de conhecimento (KB) e a interação dos agentes com o ambiente e os Analistas de ABC. As letras A, B e C presentes no diagrama permitem conexão com o fluxo mostrado na Figura 5.2 [6].

5.3.4 Descrição Funcional dos Agentes

O agente capturador de perfis suspeitos (CPS) pode ser único ou formar um grupo especializados por produto da linha de negócios de uma instituição financeira (contas correntes, câmbio, fundos de investimento, empréstimos etc.). A especialização oferece vantagens, tais como: cada agente pode ser aprimorado com as especificidades do produto e dos perfis dos clientes, que podem mudar em cada produto; outra vantagem diz respeito aos princípios de manutenibilidade e de escalabilidade, ou seja, um produto vigente pode ser descontinuado (morte de um CPS) e um novo produto pode ser criado (novo CPS) sem que isso interfira no funcionamento dos demais agentes. Dentre as especificidades dos produtos é possível citar o requerimento de informações adicionais, como por exemplo,

verificação em listas restritivas internacionais de comércio entre países (câmbio) ou listas de pessoas politicamente expostas (PEP), sobre quem a legislação exige cuidado especial.

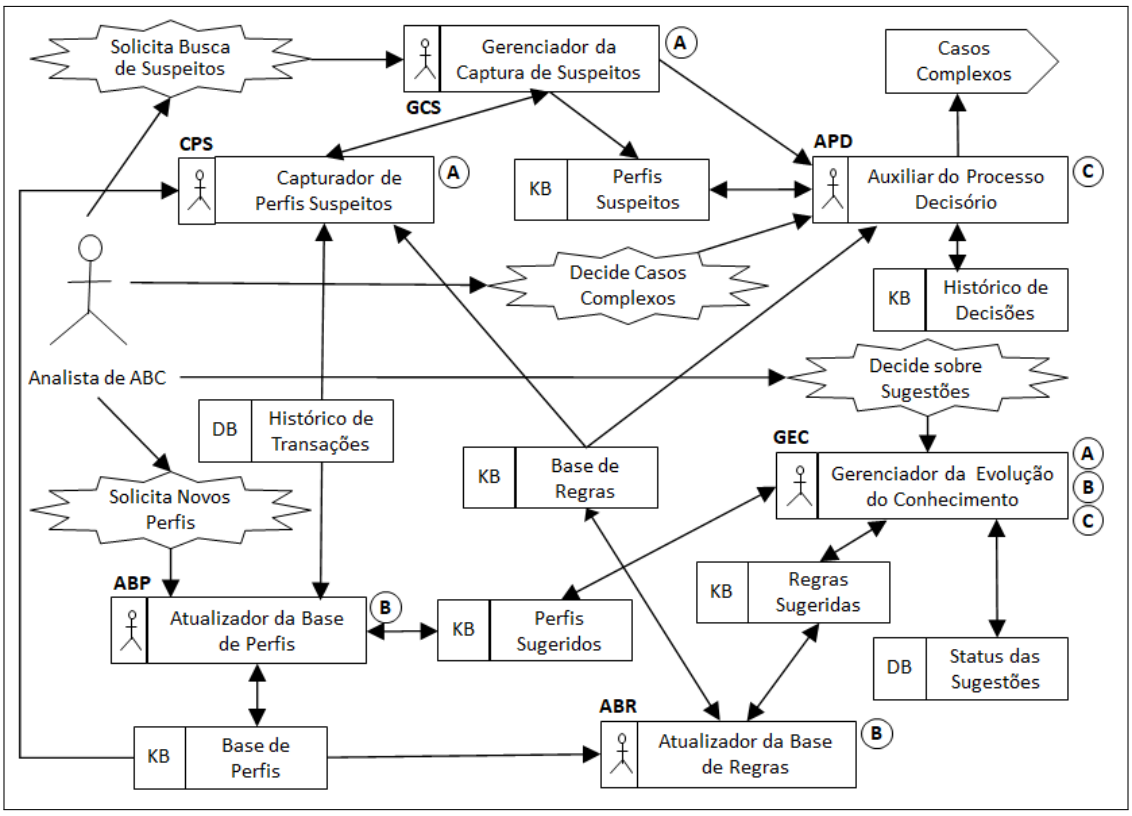


Figura 5.5: Versão Resumida do Modelo do Sistema

O banco de dados com o histórico do perfil dos clientes, visível na Figura 5.5 como DB-Histórico de Transações, refletindo o período de aprendizado, e um conjunto de regras constituem a principal base de conhecimento dos agentes. O conjunto de regras, identificado 5.5 como KB-Base de Regras é formado por:

1. DM Regras – regras geradas no processo de *data mining*, que serão usadas na revisão da classificação original dos clusters;

2. Regras Legais – regras criadas com base em regulamentos legais e diretrizes gerais, emitidas por organismos de controle, que orientam a luta contra o crime de lavagem de dinheiro;

3. Regras de Perfil – regras que refletem o conhecimento específico do banco sobre ABC e o gerenciamento de riscos de acordo com os grupos de perfis gerados.

A base geral de perfis, identificada na Figura 5.5 como KB-Base de Perfis e o histórico de decisões, identificado no modelo como KB-Histórico de Decisões, completa o conjunto de informações que serão manuseadas pelo sistema no processo de captura e análise das transações suspeitas.

Na análise das transações os agentes utilizam as regras de produção vigentes, geradas com base nos perfis dos clientes, nos normativos sobre ABC e nas normas internas da instituição financeira. Atuam em duas modalidades: busca por perfil ou busca por cliente específico. Na busca por perfil a base histórica de transações é analisada integralmente, dentro do período informado, enquanto na busca por cliente, somente as transações relacionadas ao cliente informado são analisadas.

O agente Gerenciador da Captura de Suspeitos (GCS) pode receber uma solicitação externa de análise e comandar para execução pelo Capturador de Perfis Suspeitos (CPS) especialista ou comandá-la autonomamente. Ao receber a informação de um CPS de que uma transação suspeita foi identificada, comanda uma análise na modalidade busca por cliente para os demais CPSs responsáveis por produtos que constem do perfil do referido cliente. Somente o GCS tem conhecimento de quantos e quais são os CPS existentes. Após todos os CPSs acionados terem enviado mensagens de resposta ao comando de análise, o GCS comunica ao agente Auxiliar do Processo Decisório (APD) a existência de transações suspeitas.

O agente Auxiliar do Processo Decisório (APD) é responsável por, autonomamente, aprofundar a análise e decidir sobre a suspeição do cliente, confirmando-a ou não. A utilização de conhecimento específico sobre o produto e de uma base de dados com o histórico de decisões tomadas, permitem ao agente decidir sobre a suspeição ou, quando não é possível chegar a uma decisão, sinalizar para o Analista de ABC sobre a existência de um caso complexo. Os casos sinalizados para análise humana aplicam o Índice de Suspeição no intuito de orientar o trabalho do Analista. As decisões tomadas pelo agente, bem como aquelas informadas pelo Analista, são guardadas no histórico de decisões e são utilizadas como base para criação de novas regras de perfil.

O conhecimento interno do sistema, necessário para o seu funcionamento, é formado pelas bases de perfis, normativos, de regras e pela já citada matriz de decisão. Para cada uma dessas bases de conhecimento existe um agente responsável por sua evolução.

O agente Atualizador da Base de Perfis (ABP) atua na análise do histórico de transações para geração de perfis de clientes e posterior comparação com a base de perfis existente. Este processo pode ser acionado por uma solicitação do usuário ou de forma autónoma pelo agente. Os possíveis novos perfis surgidos são sugeridos para o Analista de ABC, para análise. O agente ABP atualiza a base de perfis com os perfis que forem validados.

A base de regras de produção é um elemento chave na arquitetura do sistema, sua evolução é feita pelo agente Atualizador Base de Regras (ABR). A primeira etapa deste processo de evolução consiste na geração de novas regras, confrontadas com a base de regras vigente, baseadas nos novos perfis. O ABR sugere então essas novas regras para análise do Analista de ABC. As regras resultantes desta análise são incorporadas à base de regras que será utilizada pelos agentes decisores.

Algumas tarefas são muito importantes no sistema e dizem respeito a controlar as sugestões feitas pelos agentes para evolução das bases de conhecimento; manter interface

com o Analista de ABC no procedimento de validação, eliminação ou ampliação dessas sugestões; e manter comunicação com os demais agentes para comandar e controlar a efetivação dessas evoluções. Esta tarefa é desempenhada pelo agente Gerenciador da Evolução do Conhecimento (GEC). A interface mantida com o Analista de ABC permite que este analise e valide as sugestões além de garantir que esta tarefa é realizada somente por analistas cadastrados.

As evoluções das bases de conhecimento e o aprendizado previstos no sistema visam, primordialmente, mitigar o risco da ocorrência de falso positivo e/ou falso negativo, existente em sistema baseados unicamente em um conjunto de regras e padrões de comportamento [63, 79].

5.3.5 Interação entre os Agentes

O código e diagrama *AUML* utilizados na definição dos protocolos de comunicação utilizam quase integralmente as definições apresentadas por Winikoff [126]. A estrutura escolhida para implementar o sistema, que será discutida a seguir, usa nativamente a estrutura interna do mecanismo de tomada de decisão do modelo do Procedural Reasoning Sytem (PRS) e a KQML para estabelecer um processo de comunicação simples, porém eficiente. "Por exemplo, a KQML performativa *tell* é usada com a intenção de alterar as crenças do recetor, enquanto que *achieve* é usado com a intenção de alterar os objetivos do recetor", ou seja, através do *label* da performativa é possível identificar a intenção do remetente da mensagem, oferecendo flexibilidade de ação [23].

A cooperação entre os agentes CPS ocorre por meio da interação direta com o agente GCS, que coordena as tarefas e recebe os resultados. Somente o agente CGS conhece todos os agentes CPS existentes; além disso, ele pode identificar os produtos que o perfil suspeito usa na instituição. A figura 5.6 mostra um dos processos de interação entre agentes.

A interação entre outros agentes (APD, ABP, ABR e GEC) tem o papel de acionar em cada um desses agentes o objetivo de executar a tarefa sob sua responsabilidade. Em outras palavras, todos os agentes têm seus próprios conhecimentos específicos e têm objetivos independentes e não conflituante. Por outro lado, há muita cooperação para a consecução de um objetivo comum.

5.4 Implementação do Sistema

5.4.1 Ambiente Utilizado para Implementação

Partindo de Fisher et al. [59] e com base em Boissier et al. [20] e Shehory and Sturm [114], o *framework* JaCaMo [22] foi escolhido. O tutorial apresentado em AAMAS2015 por Baldoni et al. [10] e o ótimo posicionamento da plataforma no estudo realizado por Kravari and Bassiliades [76] reforçaram esta escolha. A estrutura é flexível na interação

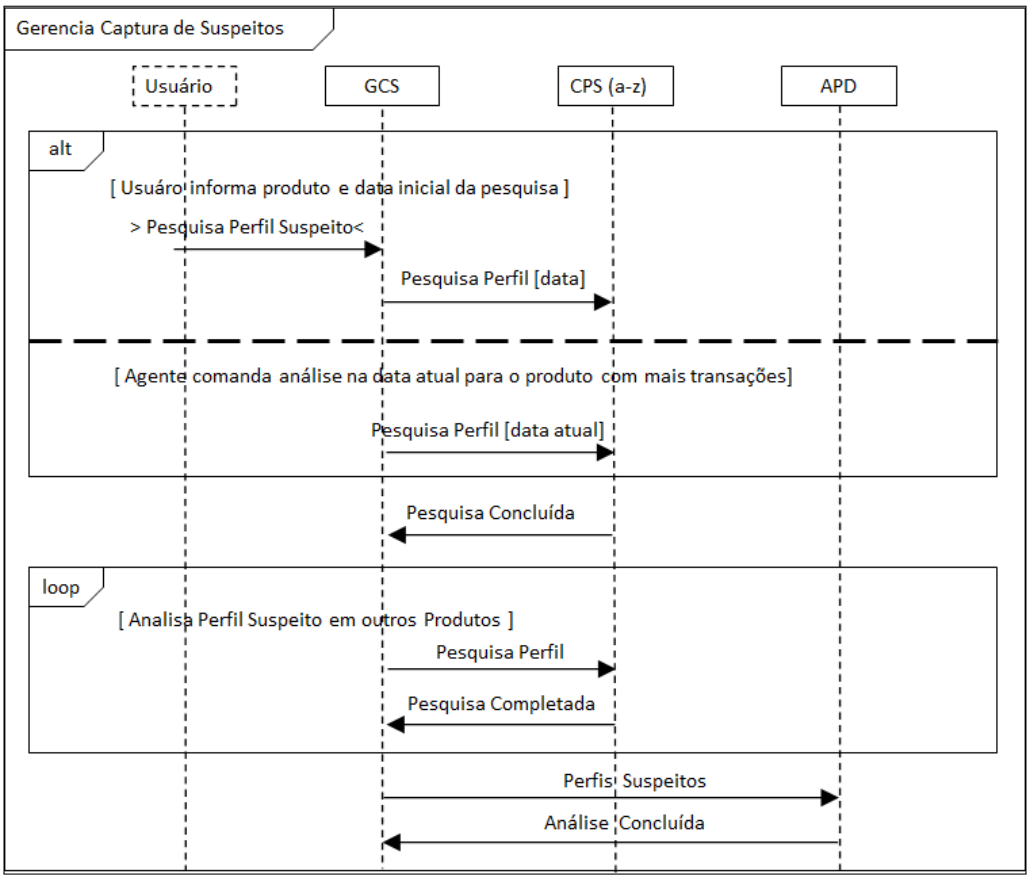


Figura 5.6: Exemplo do Processo de Interação entre Agentes

com outras linguagens de programação, principalmente java; acesso nativo às principais plataformas de banco de dados; boa usabilidade e documentação; compliance com a FIPA; licença grátis; e BDI nativo [8, 76].

O JaCaMo é baseado em três plataformas independentes: a) *Jason* para programar o nível de agentes, inspirado na arquitetura BDI; b) *CArtAgO* para programar o nível do ambiente, que é composto por um ou mais espaços de trabalho, usados para definir a topologia do ambiente; c) *Moise* para programar o nível das organizações, definindo os grupos e subgrupos sociais existentes, tarefas sociais e regras de comportamento que permitirão a consecução de objetivos sociais [22, 23].

As regras obtidas no processo de mineração de dados foram escritas como regras de Jason e usadas em um processo semelhante ao usado em Prolog. As bases de conhecimento e bancos de dados externos são acessados no MySQL usando artefactos escritos em Java para melhorar o desempenho no acesso, devido ao volume de dados. Todo conhecimento e bancos de dados externos são usados pelo sistema como crenças, representando os estados do ambiente e o conhecimento adquirido.

Os principais elementos do sistema podem ser representados no modelo BDI apresentado na Figura 5.7, diante disto, a seguir serão mostrados exemplos de como estes elementos foram implementados.

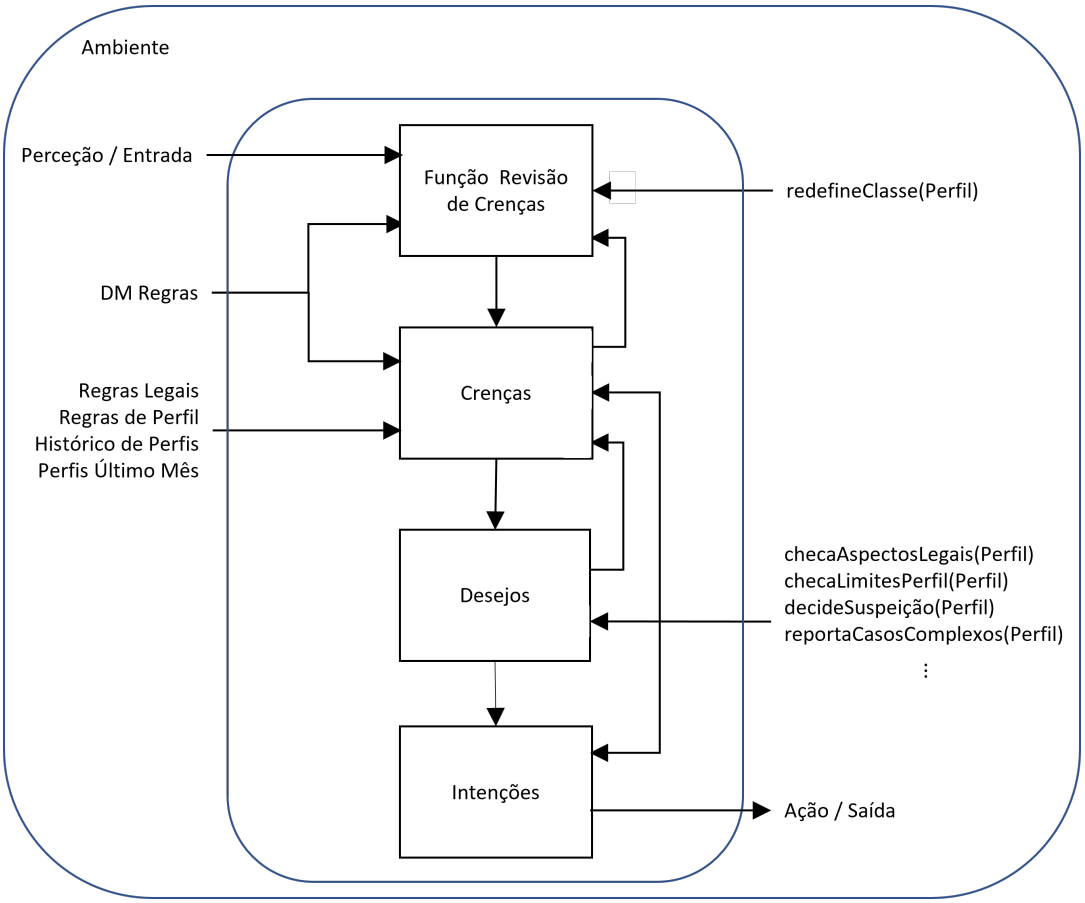


Figura 5.7: Modelo BDI do Sistema

5.4.2 Lidando com o Conhecimento sobre o Comportamento Transacional

Partindo-se do princípio que as crenças representam os conhecimentos que o agente possui acerca do ambiente que ocupa e faz parte, os perfis representam o conhecimento acumulado sobre cada cliente. O perfil concentra, de forma codificada, o histórico, o comportamento e a classificação de risco de um cliente em comparação com os demais. Este conhecimento, assim como outra grande quantidade de crenças, precisa estar totalmente presente na memória de um agente no momento da tomada de decisão.

Considerando a grande quantidade de dados (perfis) a serem analisados a cada processamento, do ponto de vista de desempenho torna-se inviável o acesso à base de dados a cada perfil a ser analisado. Dessa forma, o mais racional é que a maior quantidade possível de perfis esteja na memória de um agente. Isto leva a um contingenciamento em função da quantidade de memória física disponível no equipamento que executará a análise.

A estratégia adotada foi separar os conhecimentos entre técnico e comportamental. O conhecimento técnico concentra o que é necessário para embasar a tomada de decisão e o comportamental é aquele que permite aplicar, na sua definição básica, os conceitos do *Know Your Customer (KYC)*. O conhecimento técnico é estático à luz do ambiente atual

enquanto o comportamental depende da qualidade de perfis analisados. Por isso, o agente que detém o conhecimento sobre o comportamento transacional é limitado e precisa ter seu conhecimento substituído a cada nova quantidade de perfis a serem analisados. A quantidade de perfis que estarão sob o conhecimento de um agente depende da disponibilidade de memória física.

A solução encontrada foi a de criar agentes temporários, que atuam como auxiliares aos CPS, denominados Recuperador de Perfis para Análise (RPA) limitados à quantidade de perfis pré-definidos (Código C.1 linha 17), sendo criados e eliminados tantas vezes quantas forem necessárias para expressar todo o conhecimento dos perfis que executaram transações no período analisado. A Figura 5.8 mostra o modelo de tratamento desses agentes, salientando que a cada momento existe apenas um agente RPA em ação. O Código C.2, Linhas 3 a 17, mostra processo de criação e eliminação dos agentes RPA

No caso mostrado na Figura 5.8, a quantidade total de perfis selecionados para análise é dividida pela quantidade de perfis definida como limite, em função da restrição de memória física (250.000 / 40.000 = 6,25). A lógica então é montada para criar sete agentes RPA, sendo que seis trabalharão com 40.000 perfis e o sétimo com 10.000 perfis.

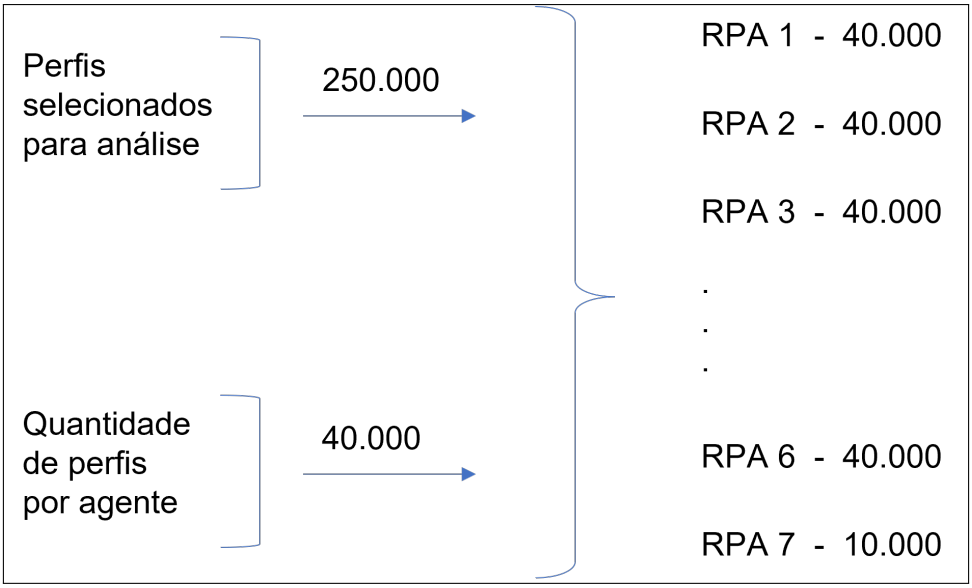


Figura 5.8: Esquema de Atuação dos Agentes RPA

5.5 Interface e Rotinas Relevantes do Sistema Multi-Agente

Com o objetivo simples de organizar, sequenciar e facilitar a utilização dos elementos definidos e descritos neste trabalho, foi elaborado um protótipo de um Sistema Multi-Agente batizado de Jano. A origem do nome vem da mitologia romana e também etrusca segundo a qual Jano (do latim Janus ou Ianus) era um deus representado por duas cabeças, buscando simbolizar os términos e os começos, o passado e o futuro, o dualismo relativo de todas as coisas. Ele seria o porteiro da corte divina, podendo olhar para lugares opostos

e vigiar as interseções dos caminhos. Jano foi a inspiração do nome do primeiro mês do ano (janeiro, do latim januarius). Ele teria inventado o dinheiro e as primeiras moedas traziam sua imagem [95, 109]. O simbolismo do nome está na capacidade do sistema de vigiar os diversos caminhos trilhados pelas tentativas de Branqueamento de Capitais.

A interação inicial do sistema com o ambiente pode ser observada na Figura 5.9 cujas principais informações estão a seguir descritas:

- **(1)** – a primeira janela controla a execução das principais ações desta versão do sistema e no destaque um são mostradas os botões de acionamento. As características especificas de cada função serão descritas ao longo deste capítulo;

- **(2)** – na secção dos agentes é mostrada a relação de agentes ativos e também os que foram eliminados ao longo do processo;

- **(3)** – neste quadro fica a relação de arquivos disponíveis resultante de processos já executados;

- **(4)** – esta janela apresenta a interação do agente GCS com o ambiente, bem como a execução das duas primeiras fases do processo de captura. As funções dos agentes envolvidos neste processos estão descritas em 5.3.4 (pág. 79);

- **(5)** – a secção chamada Ambiente Atual mostra os valores que serão utilizados como parâmetros para execução dos agentes envolvidos. Algumas dessas informações serão transformadas em crenças e outras para localização de base de dados e montagem de consultas ao banco de dados. Informações como Perfis por Agente, informa a quantidade de perfis que farão parte de cada grupo a ser analisado, necessidade imposta por uma limitação de memória comentada em 5.4.2 (pág. 84). Outra informação mostra a Margem Adicional de Risco (MAR), explicada em 4.4 (pág. 64), previamente definida. As opções Cliente, Agência, Conta e Dígito, quando informadas, permitem executar análise de um cliente específico. A informação de Movimento determina o mês e ano de referência para o processamento[3];

- **(6)** -– esta área permite modificar as informações então vigentes no ambiente, descritas no item anterior;

- **(7)** -– semelhante à janela anterior, esta secção mostra os agentes ativos e eliminados sob a gestão do agente CPS. Cada agente Recuperador de Perfis para Análise (RPA), mostrados como ativo e eliminados, processa um grupo de perfis cuja quantidade foi definida pelo número referenciado no item 5. Detalhes sobre a lógica aplicada pelo agente Recuperador de Perfis para Análise (RPA) está descrita em 5.4.2 (pág. 84);

[3]Em consequência de compromisso assumido com a Instituição Financeira fornecedora dos dados, serão mantidas em sigilo as informações que indiquem o ano de execução das transações ou permitam identificação do cliente, mesmo que parcial (código do cliente, agência e conta).

- **(8)** -– este quadro mostra o andamento e o resultado final das fases em execução e será detalhado a seguir.

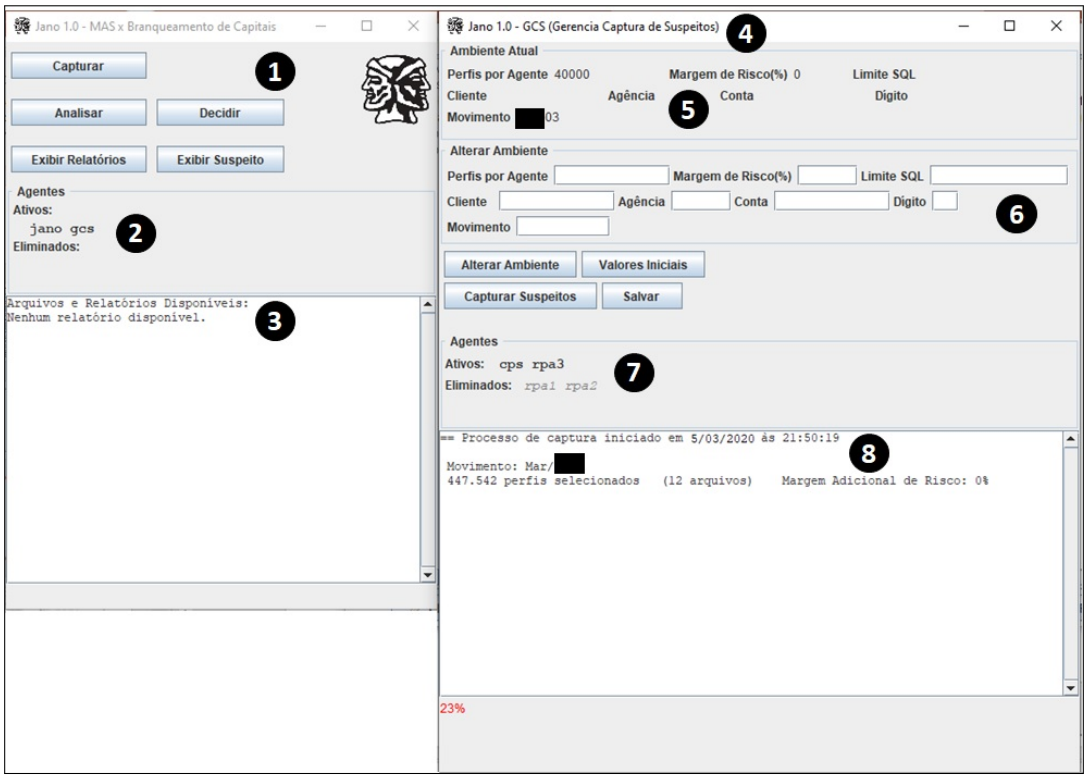


Figura 5.9: Interface Inicial do Sistema

O código-fonte das definições rotinas mais relevantes do sistema estão colocados no Apêndice C, na sequência descrita a seguir:

- **Crenças Globais** – Para um melhor funcionamento do sistema, alguns valores foram parametrizados pois representam informações externas (localização de ficheiros - código C.1 linha 14) ou aspetos gerais que auxiliam a tomada de decisão dos agentes, que não devem ficar no código, quer seja pela possibilidade de mudança, quer seja pela repetição (codificação dos perfis, Índice de Suspeição (IS), Margem Adicional de Risco (MAR) etc.). O ambiente em uso permite esta codificação num arquivo que pode ser incluído no início da execução dos agentes, conforme mostrado abaixo. Os comentários no código buscam explicar a função de cada parâmetro;

- **Reclassificação dos Perfis** – Conforme detalhado em Abordagem Conservadora em Relação ao Nível de Risco (tópico 4.4 - pág. 64) foi adotada a estratégia de reclassificar os perfis Baixa Utilização, Cliente Padrão e Risco Baixo, com o objetivo de corrigir o problema provocado quando uma regra classifica um cliente como pertencente a mais de um perfil. Quando isto ocorre o cliente é temporariamente reclassificado para o perfil com indicativo de maior risco. Inicialmente o agente CPS

obtém do agente RPA responsável todas as informações que possui sobre o perfil em análise (Código C.2, Linha 20), em seguida, a classe é redefinida e, dependendo do tipo de cliente, as regras são novamente aplicadas efetuando a reclassificação (Código C.3). Considerando que o perfil atribuído inicialmente a cada cliente é definido no sistema como uma crença, realizar esta reclassificação temporária pode representar uma revisão da crença inicial;

- **Codificação das Regras** – Um exemplo de codificação das regras obtidas pelo processo de aprendizagem é mostrado no Código C.4. As regras são codificadas de forma a permitir sua fácil localização e expressar o conhecimento adquirido.

Capítulo 6

Apresentação e Análise dos Resultados

6.1 Contexto

Neste capítulo são apresentados os resultados obtidos com a aplicação de cada uma das técnicas e teorias apresentadas anteriormente. Inicialmente são apresentadas características sobre os dados utilizados neste trabalho e os resultados em cada etapa do processo.

6.2 Sobre os Dados Envolvidos

Os dados reais utilizados neste trabalho referem-se a dois anos de movimentações, com 30,5 milhões e 35,2 milhões de transações relevantes, respetivamente, conforme mostra a Tabela 6.1. Transações relevantes são aquelas que, considerando suas características, direta ou indiretamente, podem vir a ser consideradas suspeitas de Branqueamento de Capitais (BC). O elemento básico para a definição da relevância está na informação existente na tabela denominada Históricos, cujos códigos definem a natureza ou finalidade das transações. A relevância desses códigos para a tarefa de Anti-Branqueamento de Capitais (ABC) foi validada por Analistas de ABC da Instituição Financeira que forneceu os dados e patrocinou esta pesquisa. Como exemplo é possível citar que transações originadas pela própria instituição financeira não são consideradas relevantes e foram desprezadas (tarifário, coimas, juros etc.), por outro lado, transações envolvendo levantamentos, depósitos, transferências, pagamentos etc., são consideradas relevantes.

Perfil do Cliente é uma entidade criada neste trabalho objetivando concentrar informações que permitem estabelecer o comportamento transacional de um cliente, conforme detalhado em 4.3.1 (pág. 51). Perfis, na Tabela 6.1, representa a quantidade de clientes com transações relevantes no ano em análise. Apesar da quantidade de registos do Cadastro ter apresentado um crescimento de 18%, os perfis mostraram variação de apenas 8%, o que pode ser explicado pelo número de clientes que abriram contas corrente e não realizaram movimentações no período ou referem-se a contas criadas com a finalidade exclusiva de recebimento de salários, situação descartada como passível de BC. Impor-

tante destacar que 1,7 milhão de clientes, em torno de 70%, realizaram transações nos dois anos envolvidos no estudo, razão pela qual são chamados de clientes frequentes.

Os perfis de comportamento transacional dos clientes foram gerados a partir dos dados do primeiro ano, constituindo a base de referência. A busca por transações suspeitas foi realizada em seis meses de transações relevantes do segundo ano, resultando na análise média de 2,8 milhões de transações/mês. Sobre estas transações foram selecionados em média 441,7 mil perfis/mês com comportamentos transacionais passíveis de análise, conforme mostra a Tabela 6.2.

Tabela 6.1: Números sobre os Dados Envolvidos(*)

	Clientes Ano 1		**Clientes Ano 2**		**Perfis**
Tabelas	**Total**	**Perfis**	**Total**	**Perfis**	**Frequentes**
Cadastro	5,1	2,4	6,0	2,6	1,7

	Transações Ano 1		**Transações Ano 2**	
	Total	**Relevantes**	**Total**	**Relevantes**
Movimentos	90,6	30,5	82,4	35,2

(*) Todos os valores expressos em milhões

Tabela 6.2: Perfis Selecionados para Análise

Ano 2 - Mês Referência	**Transações Relevantes**	**Perfis Selecionados**
Jan	2.734.718	444.819
Fev	2.435.385	408.302
Mar	2.918.980	447.542
Abr	2.786.800	427.415
Mai	2.833.358	447.683
Jun	3.081.400	474.484
Média	**2.798.440**	**441.708**

A decisão de analisar seis meses de dados justifica-se pelo fato da validação dos resultados obtidos, necessariamente, ter de ser realizada por analistas humanos da instituição financeira envolvida no estudo, portanto, um número muito grande de perfis sinalizados como suspeitos poderia tornar a validação inviável.

Sobre a necessidade desta avaliação por terceiros, é preciso relembrar o ambiente descrito em 2.2.3 (pág. 21), denominado neste trabalho de risco sistémico, resultante da falta de retroalimentação do processo de ABC. Os órgãos reguladores não informam a efetividade das comunicações de clientes com transações suspeitas, realizadas pelas instituições

financeiras. A instituição comunica a suspeita, mas não sabe se esta suspeita se confirmou ou não. Esta situação praticamente impede a utilização de técnicas supervisionadas.

Conforme salientou Ebberth Paula em [97], talvez a maior dificuldade no uso de técnicas não supervisionadas seja exatamente a avaliação dos resultados em relação aos objetivos propostos. A avaliação por especialistas é subjetiva e, portanto, dificilmente estarão totalmente alinhadas com os aspetos tratados pelos algoritmos. Contudo, este é o processo normal de trabalho e, nesta pesquisa, o compromisso dos Analistas de ABC é para com a instituição que financiou a pesquisa e não para com os formuladores da pesquisa. Assim sendo, entendemos como válida esta avaliação.

6.3 As Três Fases do Processo

O processo de busca por transações suspeitas foi dividido em 3 fases: a reclassificação dos perfis ou ajuste da matriz de confusão; a captura dos perfis suspeitos; e a análise dos perfis capturados. As duas primeiras fases são processadas pelos agentes Gerenciador da Captura de Suspeitos (GCS) e Capturador de Perfis Suspeitos (CPS) e representam a interação com o ambiente, a revisão de crenças e a captura de perfis suspeitos. A terceira fase é executada pelo agente Auxiliar do Processo Decisório (APD) após receber do agente Gerenciador da Captura de Suspeitos (GCS) a sinalização da existência de perfis suspeitos.

6.3.1 Fase 1 - Reclassificação dos Perfis

Um resumo do resultado da aplicação da estratégia de reclassificação dos perfis, descrita em Abordagem Conservadora em Relação ao Nível de Risco (tópico 4.4 - pág. 64), é mostrada na Figura 6.1. As principais informações do resumo estão identificadas pela numeração de um a três, que serão descritas a seguir:

- **(1)** – o primeiro número existente na linha indica a quantidade de perfis selecionados para análise, conforme mostrado previamente na Tabela 6.2. Informa também que esta quantidade de perfis foi dividida em 12 arquivos, em função do número de Perfis por RPA. Em seguida é informado o percentual utilizado na Margem Adicional de Risco (MAR);

- **(2)** – informa a quantidade de regras e versão das regras utilizadas para efetuar a reclassificação dos perfis. Regras estas obtidas no processo mineração de dados e aprendizado de máquina, descritos em 4.3 (pág. 50);

- **(3)** – indicação de que 389 perfis tinham características que eram atendidas por regras que resultavam em mais de um *cluster*. Originalmente classificados como “padrão", foram reclassificados, ficando 32 em risco médio e 357 em alto risco. Dos perfis que representam outros tipos de pessoa, 6 foram reclassificados, saindo

do perfil padrão para o perfil de alto risco. Importante salientar que com esta abordagem 391 perfis que inicialmente seriam analisados com menos rigor, por não apresentarem características de risco, foram selecionados para verificação dentro de critérios de médio e alto risco.

6.3.2 Fase 2 - Captura dos Perfis Suspeitos

Na Figura 6.1 é mostrada as informações referentes as regras utilizadas e a quantidade de perfis selecionados como suspeitos.

- **(4)** – quantidade e versão das regras utilizadas para efetuar a captura dos perfis suspeitos. Estas regras estão divididas em regras baseadas em normativos e regras baseadas nos perfis. O primeiro conjunto de regras aplica as determinações normativas que não se baseiam em limites, ou seja, estão mais próximos de exprimirem comportamento. Determinações estas definidas pelos bancos centrais do Brasil e Portugal. O segundo conjunto permite a aplicação da abordagem da MAR e incorpora regras capturadas da experiência dos Analistas de ABC;
- **(5)** – apresenta a quantidade de perfis capturados como suspeitos e que serão analisados na fase seguinte.

As regras extraídas da experiência dos Analistas de ABC, citadas no item quatro, permitem a incorporação de conhecimentos do tipo: perfis ligados a atividades identificadas como de risco são analisados como de alto risco e levam mais em consideração a quantidade de transações que os valores envolvidos; contas novas (menos de um ano) são selecionadas sempre que movimentam valores nas faixas mais altas ou efetuam grande quantidade de transações.

Considerando que a utilização da MAR é opcional e sua variação no sentido mais conservador ou mais liberal, a critério do utilizador, modifica o comportamento original do sistema, os resultados apresentados a seguir e que serão validados por analistas humanos utilização o valor da MAR sem variação, ou seja, igual a 100%.

O número de suspeitos, para um mês de transações, mostra-se exequível para a análise humana, dentro de um processo normal e oficial de análise.

6.3.3 Fase 3 - Análise dos Perfis Capturados

Esta fase do processo aciona o agente Auxiliar do Processo Decisório (APD) que busca fornecer ao Analista de ABC as informações necessárias para uma correta tomada de decisão, podendo o próprio agente decidir nos casos de menor complexidade. As principais informações fornecidas nesta fase estão mostradas na Figura 6.2, identificadas da seguinte forma:

- **(1)** – resumo com a distribuição dos suspeitos pelos grupos de perfis;

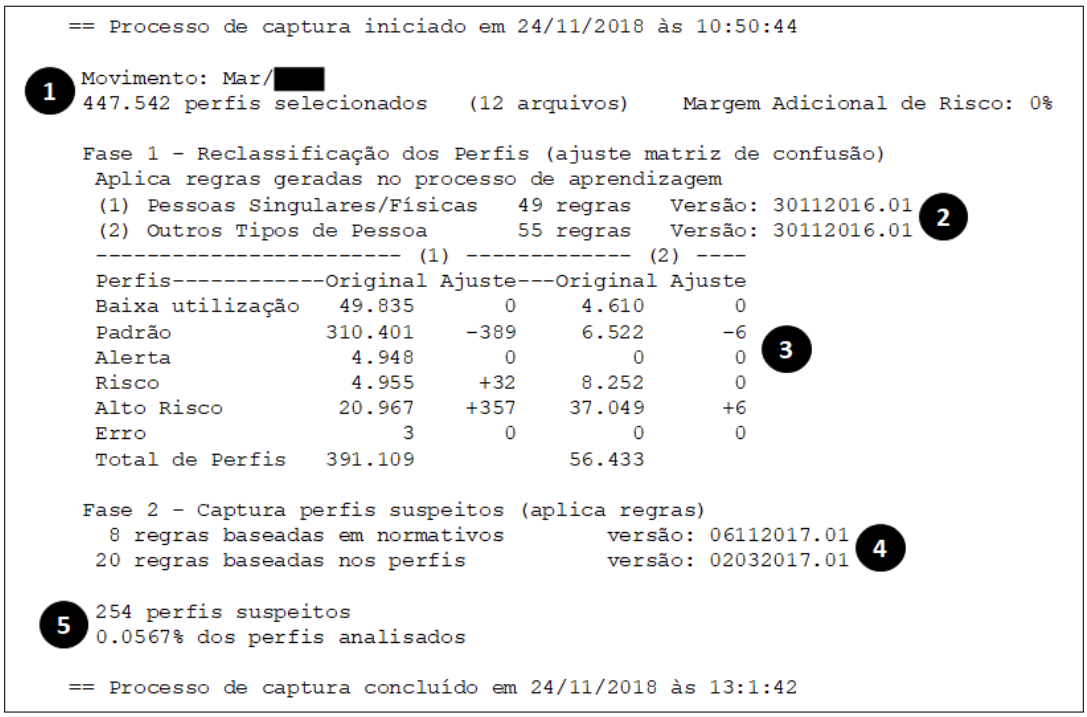


Figura 6.1: Processo - Fases 1 e 2

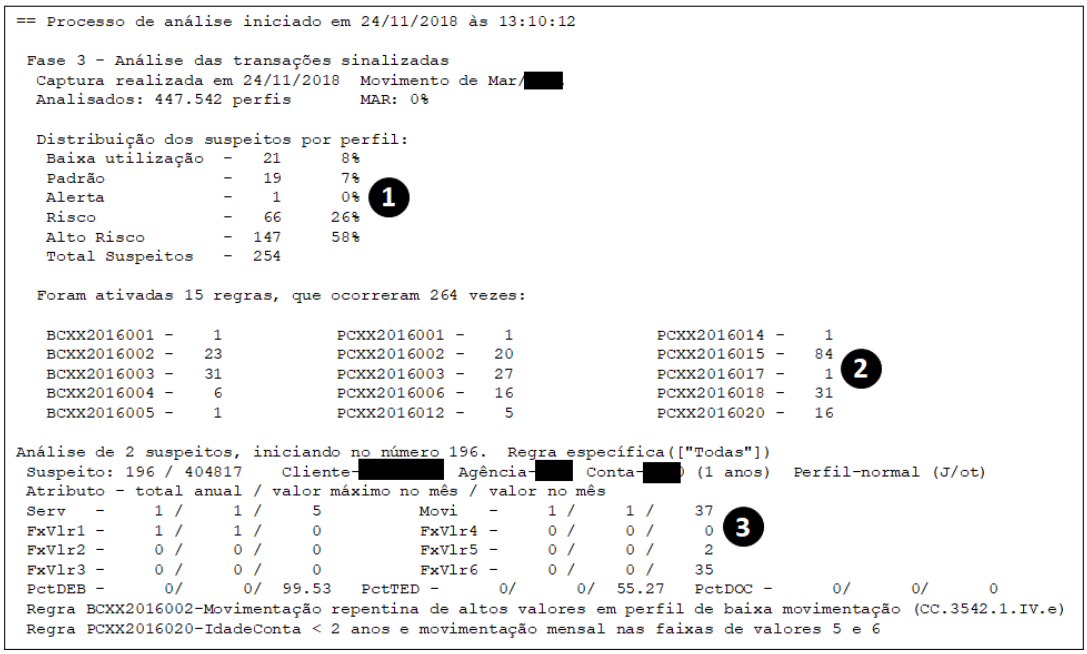
- **(2)** – apresenta o nível de utilização das regras aplicadas aos suspeitos. Assim como a informação anterior, o objetivo é auxiliar o Analista a identificar possível concentração de casos em poucas regras sinalizando a necessidade de revisão do modelo, regras ou grupos de perfis;

- **(3)** – este trecho mostra em detalhes as informações do perfil indicado como suspeito, bem como, as regras ativadas para o caso;

- **(4)** – na parte final do relatório os casos suspeitos são classificados segundo Índice de Suspeição (IS), visando oferecer uma priorização de investigação. Informações sensíveis foram omitidas. Os números de identificação do perfil suspeito são internos do sistema;

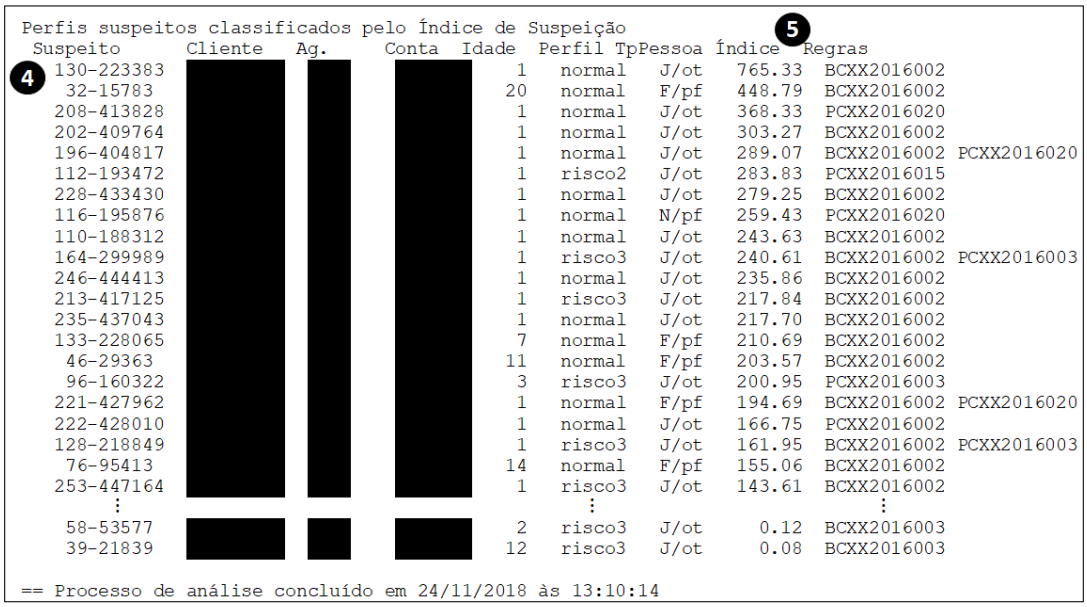
- **(5)** – mostra a aplicação do IS, que classifica os suspeitos, conforme descrito em 4.5 (pág, 67). A identificação das regras envolvidas são apresentadas para cada perfil selecionado.

A análise dessas informações precisa considerar todo o conjunto oferecido. É possível observar que a distribuição dos suspeitos por perfil, alerta para a concentração de 58% dos suspeitos no perfil de Alto Risco. Esta informação pode induzir o Analista de ABC a concentrar as investigações neste grupo de perfis. Porém, o agente também informa que os maiores IS são classificados como pertencentes ao grupo Normal, ou seja, perfis Baixa Utilização e Padrão, que foram analisados como sendo de Risco ou Alto Risco.



```
== Processo de análise iniciado em 24/11/2018 às 13:10:12

 Fase 3 - Análise das transações sinalizadas
  Captura realizada em 24/11/2018  Movimento de Mar/[redacted]
  Analisados: 447.542 perfis       MAR: 0%

  Distribuição dos suspeitos por perfil:
   Baixa utilização  -   21      8%
   Padrão            -   19      7%
   Alerta            -    1      0%
   Risco             -   66     26%
   Alto Risco        -  147     58%
   Total Suspeitos   -  254

  Foram ativadas 15 regras, que ocorreram 264 vezes:

   BCXX2016001 -    1          PCXX2016001 -    1          PCXX2016014 -    1
   BCXX2016002 -   23          PCXX2016002 -   20          PCXX2016015 -   84
   BCXX2016003 -   31          PCXX2016003 -   27          PCXX2016017 -    1
   BCXX2016004 -    6          PCXX2016006 -   16          PCXX2016018 -   31
   BCXX2016005 -    1          PCXX2016012 -    5          PCXX2016020 -   16

Análise de 2 suspeitos, iniciando no número 196.  Regra específica(["Todas"])
 Suspeito: 196 / 404817    Cliente-[redacted]  Agência-[redacted]  Conta-[redacted]  (1 anos)  Perfil-normal (J/ot)
 Atributo - total anual / valor máximo no mês / valor no mês
 Serv   -     1 /     1 /     5          Movi   -     1 /     1 /    37
 FxVlr1 -     1 /     1 /     0          FxVlr4 -     0 /     0 /     0
 FxVlr2 -     0 /     0 /     0          FxVlr5 -     0 /     0 /     2
 FxVlr3 -     0 /     0 /     0          FxVlr6 -     0 /     0 /    35
 PctDEB -      0/      0/  99.53   PctTED -      0/      0/  55.27   PctDOC -      0/      0/      0
 Regra BCXX2016002-Movimentação repentina de altos valores em perfil de baixa movimentação (CC.3542.1.IV.e)
 Regra PCXX2016020-IdadeConta < 2 anos e movimentação mensal nas faixas de valores 5 e 6
```

(a) Detalhes sobre os Perfis Capturados



```
Perfis suspeitos classificados pelo Índice de Suspeição
  Suspeito      Cliente    Ag.     Conta  Idade  Perfil TpPessoa Índice  Regras
 130-223383                                  1   normal   J/ot   765.33  BCXX2016002
  32-15783                                  20   normal   F/pf   448.79  BCXX2016002
 208-413828                                  1   normal   J/ot   368.33  PCXX2016020
 202-409764                                  1   normal   J/ot   303.27  BCXX2016002
 196-404817                                  1   normal   J/ot   289.07  BCXX2016002 PCXX2016020
 112-193472                                  1   risco2   J/ot   283.83  PCXX2016015
 228-433430                                  1   normal   J/ot   279.25  BCXX2016002
 116-195876                                  1   normal   N/pf   259.43  PCXX2016020
 110-188312                                  1   normal   J/ot   243.63  BCXX2016002
 164-299989                                  1   risco3   J/ot   240.61  BCXX2016002 PCXX2016003
 246-444413                                  1   normal   J/ot   235.86  BCXX2016002
 213-417125                                  1   risco3   J/ot   217.84  BCXX2016002
 235-437043                                  1   normal   J/ot   217.70  BCXX2016002
 133-228065                                  7   normal   F/pf   210.69  BCXX2016002
  46-29363                                  11   normal   F/pf   203.57  BCXX2016002
  96-160322                                  3   risco3   J/ot   200.95  PCXX2016003
 221-427962                                  1   normal   F/pf   194.69  BCXX2016002 PCXX2016020
 222-428010                                  1   normal   J/ot   166.75  PCXX2016002
 128-218849                                  1   risco3   J/ot   161.95  BCXX2016002 PCXX2016003
  76-95413                                  14   normal   F/pf   155.06  BCXX2016002
 253-447164                                  1   risco3   J/ot   143.61  BCXX2016002
      ⋮                                      ⋮                          ⋮
  58-53577                                   2   risco3   J/ot     0.12  BCXX2016003
  39-21839                                  12   risco3   J/ot     0.08  BCXX2016003

== Processo de análise concluído em 24/11/2018 às 13:10:14
```

(b) Classificação pelo Índice de Suspeição

Figura 6.2: Processo - Fase 3

As regras resultantes do processo Seleciona e Unifica Regras, descrito em 4.3.3 (pág. 61), foram codificadas e armazenadas numa base de conhecimento do sistema. Esta codificação das regras buscou permitir identificar sua origem e época de criação, por meio do formato OOOOAAAANN, onde: OOOO significa a origem da regra, AAAA é o ano de criação e NN um sequencial por ano. A Tabela 6.3 detalha os nomes utilizados.

Importante também ressaltar que o texto das regras indicam a relevância das suspeitas, como por exemplo a regra legal BCXX2016002, originada no normativo do Banco Central Brasileiro que sinaliza uma repentina mudança no comportamento transacional do perfil,

Tabela 6.3: Formação do Nome das Regras

Origem	Ano	Sequencial	Descrição das Regras
DMPF	2016	01 a 49	Geradas no Data Mining para pessoas singulares
DMOT	2016	01 a 55	Geradas no Data Mining outros tipos de pessoa
PCXX	2016	01 a 20	Criadas com base nos perfis de clientes
BCXX	2016	01 a 08	Refletem os normativos dos bancos reguladores

em função de movimentação de altos valores quando regularmente este perfil tem baixa movimentação. Esta é uma das situações mais corriqueiras de BC, podendo indicar o início de processo de branqueamento ou a utilização da conta por terceiros na tentativa de esconder o real usuário. A regra PCXX2016020 agrava ainda mais a situação pois ela indica tratar-se de uma conta aberta recentemente e que os valores transacionados estão nas faixas mais altas definidas pelo sistema. O IS 289.07 capturou bem o alto risco desta sinalização, conforme pode ser observado na Figura 6.2. As regras citadas estão apresentadas na Figura 6.3.

```
//-------------------------------------------------------------------------------------------------------------
regraBC(suspeito, StRisco, "BCXX2016002")
     :- qtdeMovimento(QtdeMovi) & faixaVlr5Mes(FxVlr5Mes) & faixaVlr6Mes(FxVlr6Mes) &
          minMovMesFxRisco(MinMovMesFxRisco) & QtdeMovi < MinMovMesFxRisco &
          (FxVlr5Mes + FxVlr6Mes) > (MinMovMesFxRisco).
regraBC("BCXX2016002", Desc)
     :- Desc = "Movimentação repentina de altos valores em perfil de baixa movimentação (CC.3542.1.IV.e)".

//-------------------------------------------------------------------------------------------------------------
regraPC(suspeito, Criterio, "PCXX2016020")
     :- idadeConta(IdadeConta) & IdadeConta < 2 & qtdeMoviMes(QtdeMoviMes) &
          faixaVlr5Mes(FxVlr5Mes) & faixaVlr6Mes(FxVlr6Mes) &
          calcRisco(QtdeMoviMes, VlrSup, VlrInf) & (FxVlr5Mes+FxVlr6Mes) >= VlrInf.
regraPC("PCXX2016020", Desc)
     :- Desc = "Idade da Conta < 2 anos e movimentação mensal nas faixas de valores mais altos".
```

Figura 6.3: Estrutura das Regras Utilizadas

É preciso considerar que a análise do caso e a aplicação das técnicas de *Know Your Customer (KYC)* pelos gestores da agência envolvida pode resultar numa explicação aceitável para o caso e esta decisão está a cargo do Analista de ABC. Apesar desta situação, conceptualmente, ter características de falso positivo, ela escapa do padrão pois não está passível de correção, casos semelhantes devem continuar a ser notificados e somente a investigação mais aprofundada será capaz de descartá-la.

6.4 Geração da Base de Suspeitos para Teste de Qualidade

Conforme comentado anteriormente, para atestar o nível de qualidade das sinalizações do sistema, uma quantidade de casos foi submetida aos Analistas de ABC da instituição financeira participante deste estudo. Também foi dito que os Analistas não estavam totalmente à disposição deste trabalho, sendo prioritárias as suas tarefas diárias. Portanto,

seria necessária a adoção de uma estratégia para realização desses testes.

Qualificar os dados obtidos surge como a primeira tarefa a ser realizada e para isso a análise de seis meses de dados precisava ser efetuada. As Tabelas 6.4 e 6.5 mostram o conjunto de informações desde os perfis selecionados para análise até os perfis considerados suspeitos. Uma média de 441,7 mil perfis com transações relevantes para BC foram selecionados para análise. Do total de 2,6 milhões de perfis selecionados, 2.390 tiveram sua categoria ajustada para perfis de risco e de alto risco. Deste total 1.343 foram considerados perfis suspeitos e foi sobre este quantitativo que foi elaborada a estratégia para indicação dos perfis que seriam investigados pelos Analistas, considerando que o volume total era inviável de ser analisado sem interferir nos trabalhos rotineiros da instituição.

Tabela 6.4: Perfis Reclassificados por Tipo de Risco

Ano 2 - Mês Referência	**Perfis Selecionados**	**Perfis Ajustados**		**Perfis Recetores**	
				Risco	**Alto Risco**
Jan	444.819	418	0,09%	39	379
Fev	408.302	406	0,10%	33	373
Mar	447.542	395	0,09%	32	363
Abr	427.415	392	0,09%	31	361
Mai	447.683	391	0,09%	37	354
Jun	474.484	388	0,08%	28	360
Total	**2.650.245**	**2.390**	**0,54%**	**200**	**2.190**

Tabela 6.5: Perfis Suspeitos

Ano 2 - Mês Referência	**Perfis Selecionados**	**Perfis Suspeitos**	
Jan	444.819	182	0,04%
Fev	408.302	148	0,04%
Mar	447.542	254	0,06%
Abr	427.415	216	0,05%
Mai	447.683	246	0,05%
Jun	474.484	297	0,06%
Total	**2.650.245**	**1.343**	**0,05%**

Complementando as informações sobre os perfis sinalizados como suspeitos, a Tabela 6.6 mostra a distribuição dos suspeitos pelo nível de risco, ressaltando que os perfis de risco e alto risco concentram quase 80% das sinalizações.

A indicação constante de um mesmo perfil como suspeito é sempre algo que preocupa os Analistas e nem sempre é percebido, principalmente se a ocorrência não acontecer de

Tabela 6.6: Perfis Suspeitos por Tipo de Risco

Perfis por	**Ano 2 - Mês de Referência**											
Tipo de Risco	**Jan**		**Fev**		**Mar**		**Abr**		**Mai**		**Jun**	
Baixa Utilização	8	4%	9	6%	21	8%	13	6%	13	5%	15	5%
Padrão	32	18%	22	15%	19	7%	20	9%	19	8%	23	8%
Alerta	1	1%	0	0%	1	0%	0	0%	1	0%	0	0%
Risco	47	26%	42	28%	66	26%	58	27%	67	27%	76	26%
Alto Risco	94	52%	75	51%	147	58%	125	58%	146	59%	183	62%
Total	**182**		**148**		**254**		**216**		**246**		**297**	

forma sequenciada. Utilizando esta situação, a estratégia adotada para selecionar perfis para serem submetidos aos Analistas da instituição financeira foi, em primeiro lugar identificar esta contumácia e em seguida estabelecer uma linha de corte na quantidade dos perfis selecionados somente em um dos meses. Considerando os seis meses analisados, que perfis estiveram presentes mais vezes nestes conjuntos? A Figura 6.4 mostra os quantitativos envolvidos e os selecionados para verificação. Os perfis que foram indicados como suspeitos repetidamente de dois a seis meses foram todos selecionados.

Conforme definido na Equação 4.3 o Índice de Suspeição (IS) varia em função dos valores dos atributos, assim sendo, não é possível determinar uma escala de possíveis valores, um ponto ótimo de corte ou de sinalização de risco. Apenas é possível assegurar que o risco é diretamente proporcional ao valor absoluto do IS, ou seja, quanto maior o valor maior o risco e, consequentemente, a possibilidade de confirmação da suspeição. Neste caso em estudo, a linha de corte foi aplicada nos perfis com IS inferior a 25 pontos. Ressaltando que o maior IS entre os perfis sinalizados, com ocorrência em somente um mês foi 439,95 e maior IS de todo conjunto selecionado para teste, com ocorrência em todos os seis meses, foi 2093,47. Esta linha de corte considerou: a representatividade do risco comparado com o comportamento observado do perfil, 25 pontos acima do padrão observado; a representatividade do nível de risco 25 frente aos valores máximos de IS citados anteriormente, ou seja, ele representa somente 1,2% o maior risco observado; e, principalmente, a capacidade de análise da instituição financeira. O total resultante foi de 189 perfis selecionados para verificação humana.

6.5 Resultado da Análise Realizada pelos Analistas Humanos

O processo de verificação dos casos suspeitos acontece com a distribuição destes casos entre os Analistas de ABC, que executam esta tarefa diariamente. Com o propósito de não interferir no dia a dia de trabalho do setor, a tarefa foi realizada por dois analistas durante três meses.

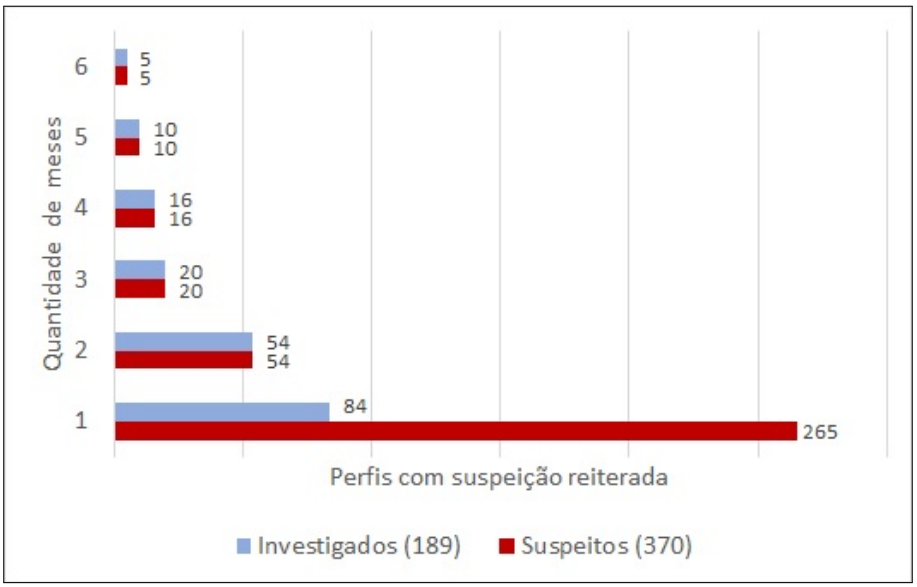


Figura 6.4: Quantidade de Perfis Selecionados para Investigação

O processo de análise, resumidamente, consiste em identificar e analisar detalhes das transações realizadas pelo perfil, tais como: dados cadastrais, natureza da transação, valores envolvidos, origem e possível destino do recurso, comparação com outras transações realizadas no período, informações publicadas sobre o cliente etc. Quando estas e outras informações analisadas não são suficientes para uma tomada de decisão, o analista demanda a agência onde o perfil mantém relacionamento solicitando informações adicionais. Caso as dúvidas persistam o analista solicita uma visita gerencial às instalações do cliente, no caso de pessoa jurídica, ou demanda atualização cadastral no caso de pessoa física.

As poucas coincidências na indicação de suspeitos demonstra a mudança de abordagem, saindo da forte parametrização de limites para análise de comportamento transacional. Consequentemente, o resultado inclui apenas algumas das suspeitas efetivamente identificadas pelo sistema em utilização pela instituição financeira e posteriormente comunicadas aos órgãos externos de controle.

A Tabela 6.7 detalha o resultado final da verificação humana. "Não Suspeito" significa que os perfis relacionados apresentam algum tipo de justificativa para a suspeição e os critérios de seleção do sistema podem carecer de análise e ajuste nas regras. Uma das situações mais comuns de não suspeição são perfis de agricultores que realizam grande quantidade de transações somente em período de colheita de safras, ficando o resto do ano com baixa movimentação. Este é o motivo para a indicação de possibilidade de ajuste, considerando que, neste caso não existe erro na sinalização.

A situação de "Suspeito, mas necessita investigar e aprofundar a análise" indica que a sinalização está correta mas com as informações obtidas pelo Analista não permite uma conclusão sobre o caso, havendo, portanto, a necessidade de novas diligências, na maioria dos casos a serem realizadas pelas agências. Geralmente são perfis cuja situação pode requerer uma visita técnica para comprovação de informações e de funcionamento

Tabela 6.7: Resultado Final da Verificação dos Perfis Suspeitos

Resultado da Análise Humana	Suspeitos Agrupados por Classificação de Risco				Sinalizados por Outros Sistemas	
	Baixa Utilização e Cliente Padrão	Baixo Risco e Médio Risco	Alto Risco	Total	Sim	Não
Não Suspeitos	35	21	44	100	81	19
Por Investigar*	8	12	23	43	26	17
Suspeições Confirmadas	13	7	26	46	11	35
Total	**56**	**40**	**93**	**189**	**118**	**71**

*requer mais informações e investigação mais detalhada, envolvendo a agência do cliente

do empreendimento. É possível observar o contributo do resultado na quantidade de casos novos, não sinalizados pelos sistemas da instituição financeira.

A "Suspeição confirmada" foram indicações que não deixaram dúvidas quanto a suspeição. Este é o status onde se apresenta o maior contributo deste trabalho, posto que as sinalizações novas representam o triplo das sinalizações efetivas reportadas pelos sistemas utilizados pela instituição financeira. Importante salientar que estes perfis deveriam ter sido reportados para os órgãos externos de controle e não foram, podendo dentre eles existirem reais praticantes do crime de Branqueamento de Capitais (BC).

Importante ressaltar que 76% dos perfis totalmente confirmados como suspeitos não foram reportados ao órgão regulador na época da sua ocorrência, porque nenhum sistema em utilização na instituição financeira sinalizou-os como suspeitos.

6.6 Variações na Utilização da Margem Adicional de Risco (MAR)

Em 4.4 (pág. 64) o conceito da Margem Adicional de Risco (MAR) foi descrito e explicado. Com o propósito de verificar sua atuação no resultado do sistema, foi realizada simulação utilizando três meses dos dados em dois cenários. Primeiro num cenário mais liberal onde foi aplicado o valor de 110%, significando que todos os limites e valores encontrados pelo sistema foram alargados em 10%, ou seja, uma tolerância de 10% na indicação de perfis suspeitos, onde se espera a indicação de uma quantidade menor de suspeitos. Um situação hipotética para esta situação seria a necessidade de gerar menos perfis suspeitos, pelos mais diversos motivos, por exemplo, falta de analistas para realizar as verificações.

Outro cenário de simulação foi mais restritivo, ou seja, um MAR de 90%, significando que os indicadores de suspeição foram todos reduzidos em 10%, onde uma maior quantidade de perfis suspeitos deve ser gerada. Para este caso os motivos também podem ser diversos, dentre eles, um teste para posterior ajuste das regras e do ambiente.

A Figura 6.5 mostra o resultado da aplicação da MAR em três meses dos dados utiliza-

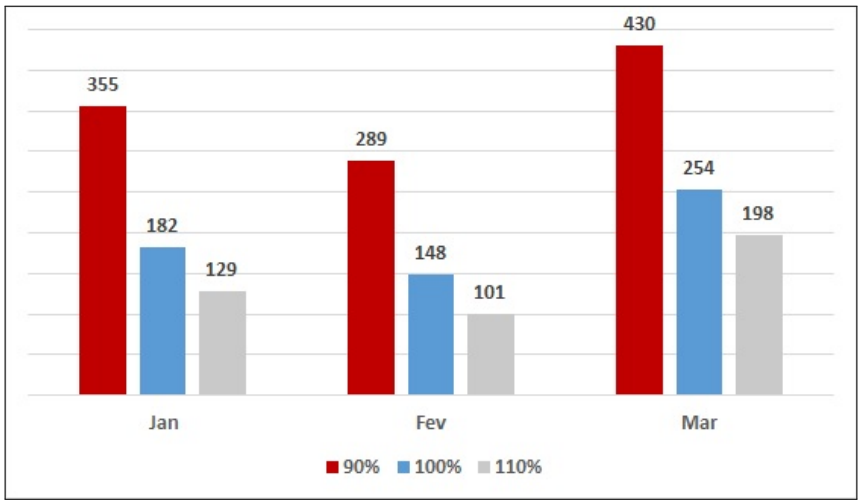


Figura 6.5: Perfis Suspeitos após Aplicação da MAR - Total em 3 meses

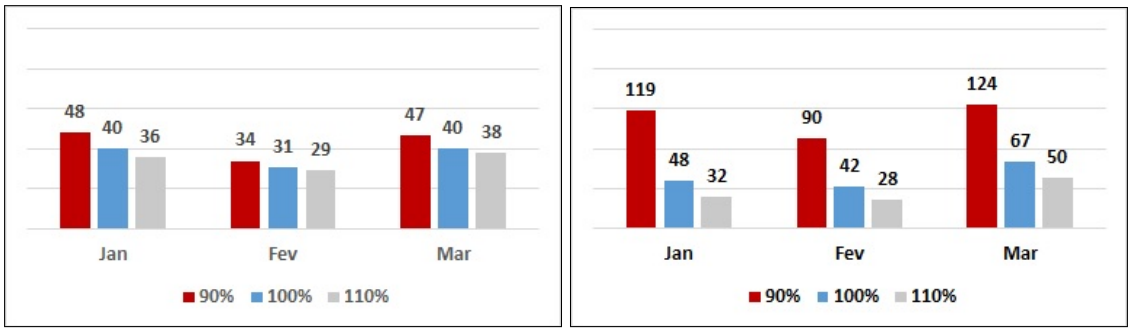
dos e, conforme esperado, demonstra a ampliação (90%) e redução (110%) na quantidade de indicação de perfis suspeitos. Com este gráfico também é possível inferir que as regras utilizadas apresentam bom nível de ajuste, uma vez que o efeito sobre o percentual mais conservador é maior que sobre o percentual mais liberal. Verificando estes dados sobre a perspetiva dos perfis, o conjunto da Figura 6.6 exibe que a MAR flexibiliza o sistema preservando seu objetivo, considerando que o efeito da MAR é maior sobre os perfis de Risco e Alto Risco quando aplicada de forma conservadora (90%) e é menor em todos os perfis quando é utilizada de forma mais liberal (110%).

6.7 Comparação com os Sistemas em Uso na Instituição

Uma forma adicional de avaliar o resultado obtido é comparando-o com os sistemas em uso. Esta avaliação pode ser feita de duas formas: comparando os valores absolutos, nomeadamente para as suspeições confirmadas; e usando métricas de avaliação de classificadores como F1-score (combinando precisão e sensibilidade) e o coeficiente de correlação de Matthews(MCC - *Matthews Correlation Coefficient*).

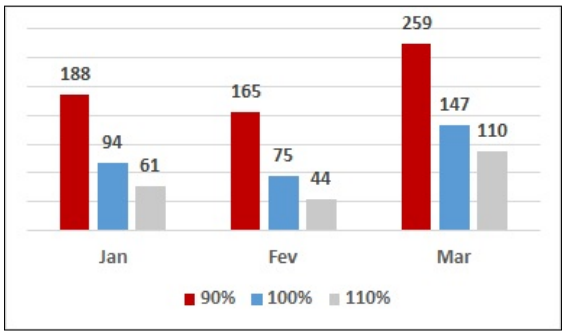
Os resultados apresentados na Tabela 6.7, e discutidos anteriormente, apresentam um claro entendimento de que, em relação aos valores absolutos, há uma melhora com a nova abordagem. A Figura 6.7 apresenta uma visão geral deste resultado, permitindo uma comparação dos resultados dos sistemas, considerando o número de perfis em cada categoria. A principal conclusão aqui, e a mais importante do ponto de vista da instituição financeira, é que o número de casos de suspeição tanto "confirmadas" quanto "por investigar" representam resultados muito melhores, com um aumento insignificante no número de falsos positivos ("não suspeitos" confirmados) [6].

Buscando ampliar e consolidar esta análise, foi efetuado o cálculo das medidas F1-score e MCC. A escolha dessas duas métricas foi motivada por F1-score ser bastante utilizada na avaliação de sistemas de deteção e por fazer uma média harmónica da precisão



(a) Perfis Normais

(b) Perfis de Risco



(c) Perfis de Alto Risco

Figura 6.6: Resultado da Aplicação da MAR por Tipo de Perfis

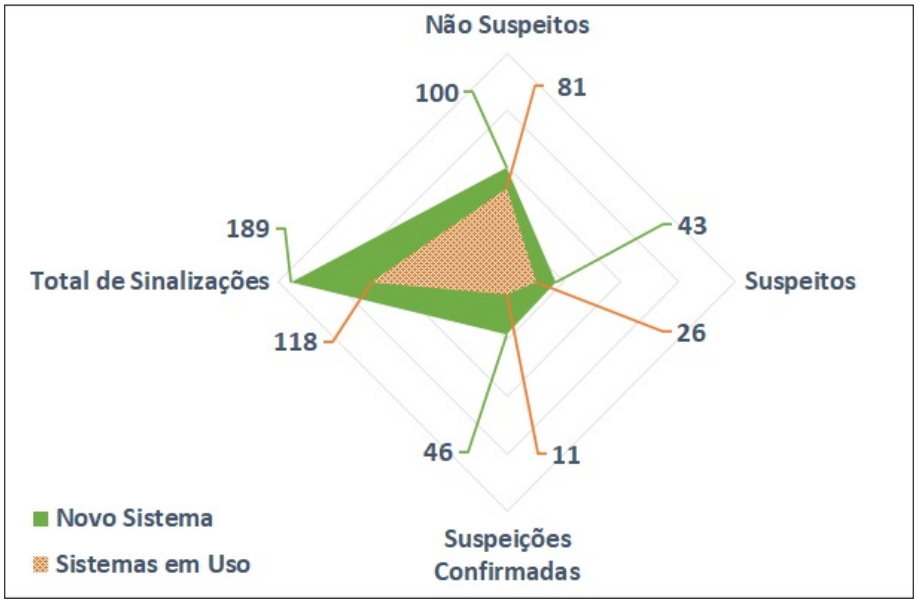


Figura 6.7: Comparação com Outros Sistemas

e sensibilidade, onde ambas precisam ser altas para que a F1-score também seja. A MCC foi escolhida por sua fórmula usar todas as possibilidades de um problema de classificação (F1-score não utiliza *verdadeiros negativos*) e por seu resultado ser de fácil interpretação, devido a normalização entre -1 (pior) e 1 (melhor). Para esta utilização foi definido o seguinte:

- São *verdadeiros positivos*, os casos com "suspeição confirmada".
- São *falsos positivos*, os casos confirmados de "não suspeito".
- Não é possível ter informações exatas sobre verdadeiros/falsos negativos, pois estes estão incluídos no conjunto de todos os casos que não foram selecionados para inspeção humana. Mas é possível fazer uma estimativa, conforme explicado a seguir.

Para execução dos cálculos, foi considerado o total de 474.484 perfis candidatos, e uma estimativa de 0,002% de *falsos negativos*[1].

Como há alguma incerteza quanto ao resultado final dos casos "por investigar", a opção foi por apresentar os resultados para F1-score e MCC usando um intervalo de valores possíveis. Assim, o limite inferior para os intervalos corresponde a considerar todos os casos sinalizados como "falsos positivos", e o limite superior corresponde a considerar todos como "verdadeiros positivos".

Os resultados obtidos nas métricas F1-score [6] e MCC são apresentados na Tabela 6.8, considerando o pior (limite inferior) e o melhor (limite superior) valor para os casos sinalizados.

Tabela 6.8: Valores das Métricas F1-score e MCC para os Sistemas Anteriores e o Novo

	Limite Inferior		**Limite Superior**		**Valor Central**	
Sistemas	**F1-Score**	**MCC**	**F1-Score**	**MCC**	**F1-Score**	**MCC**
Anteriores	0,16	0,23	0,45	0,50	0,31	0,37
Novo	0,38	0,45	0,62	0,65	0,50	0,55

É importante observar que a hipótese referente à estimativa acima é apenas uma forma de viabilizar os cálculos da medida, porém, a conclusão final (que o novo sistema melhorou os resultados) é obtida com qualquer hipótese justa (Por justa entenda-se assumir proporções, estimativas, idênticas para ambos os sistemas).

[1] Valores obtidos com base nos resultados informados nas tabelas 6.5 e 6.6

Capítulo 7

Conclusões e Trabalhos Futuros

Como capítulo final desta tese aqui serão reapresentados e discutidos tópicos referentes às questões de pesquisa, os objetivos propostos e as contribuições do trabalho desenvolvido. Ao final são apresentadas sugestões de pesquisa a serem exploradas no futuro. Nos Capítulos 2 e 3 foram inseridos comentários sobre publicações e conceitos, identificados como Nota, que serão referenciadas neste capítulo.

7.1 Conclusões e Discussões

A pesquisa desenvolvida neste trabalho situa-se na componente Informática (agentes, aprendizagem) com atuação no combate ao crime de Branqueamento de Capitais (BC) e buscou apresentar propostas reais que podem ser utilizadas pelas instituições financeiras, visando auxiliá-las no processo de Anti-Branqueamento de Capitais (ABC). Referido processo consiste em analisar todas as transações efetuadas pelos clientes dessas instituições financeiras, realizando comunicação aos órgãos competentes sobre os clientes que realizaram transações entendidas como suspeitas de BC. A falta de retroalimentação sobre a confirmação do crime dentre os clientes informados introduz o que foi definido neste trabalho com erro sistémico, além de impedir uma abordagem com maior precisão no aprendizado automático.

Diante deste cenário foram expostas questões e hipóteses no início do trabalho, para as quais são apresentadas a seguir respostas e comentários:

1. *Os algoritmos consagrados de aprendizagem indutiva não-supervisionada e a aplicação de regras por meio de agentes inteligentes são técnicas suficientes para detetar com eficácia casos suspeitos de BC, em grandes volumes de dados? se não, como torná-las?*

   Nos Capítulos 2 e 3 foram apresentadas várias publicações com propostas de novos algoritmos ou de melhoria em algoritmos existentes. Conforme comentado nas Notas colocadas em cada uma dessas publicações, elas apresentam uma característica em comum que é a não aplicação da solução proposta em grande volume de

dados. Os resultados descritos eram oriundos de simulação ou de aplicação sobre baixo volume de dados, permitindo dúvidas quanto ao desempenho (Notas E, F e G - secção 3.1.4 - [74, 119, 80]). Outras propostas, na sua totalidade ou em parte, são mais apropriadas para outros domínios, como auditoria e combate a fraude interna (Notas H e R - secções 3.1.4 e 3.2.5 - [84, 40]). Esta pesquisa demonstrou que a utilização de algoritmos de domínio geral como K-means para *clustering*, PART e J48 para geração de regras de produção são adequados para manuseio de grandes volumes de dados e identificação de transações suspeitas.

2. *Agentes inteligentes podem ser desenhados para incorporar parte da experiência dos Analistas de ABC e auxiliá-los no processo de tomada de decisão sobre os casos suspeitos de BC?*

   A utilização de agentes inteligentes na arquitetura Belief-Desire-Intention (BDI), descrito na secção 5.4, ofereceu flexibilidade para a adoção de técnicas de tomadas de decisão, notadamente na definição de crenças e na possibilidade de reclassificação dos perfis, que, temporariamente, modifica algumas crenças iniciais (secções 4.4 e 5.5). Este processo de reclassificação é fundamental para correção de situação onde o aprendizado automático identifica o cliente como pertencente a dois grupos de risco, neste caso, a crença é temporariamente revista colocando o cliente no grupos de maior risco. O funcionamento dos agentes manuseando uma base de regras oferece à solução implementada um nível adequado de extensibilidade pois além das regras oriundas do aprendizado automático permite a incorporação de regras extraídas da experiência dos analistas de ABC (secção 5.3.4).

3. *O processo de ABC pode ser desenhado de forma genérica, permitindo a definição de um sistema multi-agente que identifique e decida sobre casos suspeitos de BC?*

   É natural que cada instituição financeira acabe criando seu próprio processo de ABC, talvez isto explique que boa parte dos trabalhos publicados proponham algoritmos e não avancem no processo. Neste trabalho as definições dos órgãos reguladores internacionais, com ênfase nas regras do Banco de Portugal e do Banco Central do Brasil, foram estudadas e permitiu o desenho de fluxo genérico para o processo de ABC, que serviu de base o desenvolvimento da solução. As técnicas utilizadas permitem que cada instituição realize as adaptações necessárias sem necessitar refazer todo o fluxo (secção 5.2).

Após as pesquisas e implementações realizadas visando responder as questões acima, é possível afirmar que técnicas de *data mining* e agentes inteligentes são ótimas ferramentas para estruturar soluções cujo objetivo seja o combate ao BC. No entanto, é necessário que estas técnicas sejam aplicadas buscando sobre um processo estruturado ao invés de soluções pontuais que não atacam o problema central que é o volume sempre crescente de clientes sinalizados como suspeitos. As técnicas informáticas nem sempre são suficientes

para encontrar a solução deseja, é preciso conhecimento profundo do problema e adoção de mecanismos adicionais, por vezes simples, oferecer robustez ao produto final.

7.2 Principais Contribuições

O estudo dos trabalhos relacionados com soluções para o processo de Anti-Branqueamento de Capitais (ABC) permitiu a perceção da existência de lacunas quando defrontadas com dos problemas identificados no processo. Assim sendo, este trabalho apresenta solução para algumas dessas lacunas, resultando nas seguintes contribuições:

1. *Mudança de paradigma na abordagem genérica de combate ao crime de Branqueamento de Capitais (BC).*

   Este trabalho apresentou soluções que foram além da sinalização de transações suspeitas, fornecendo informações visando auxiliar o Analista de ABC na tomada de decisão sobre a suspeição ou não do cliente. Um dos grandes problemas identificados no estudo inicial é o aumento do volume de transações efetuadas e, consequentemente, da quantidade de transações suspeitas, sendo que o número Analistas não acompanha este crescimento. Muitos dos trabalhos publicados cumprem a função de sinalização de suspeições, sem apurar a efetividade dessas sinalizações, situação que pode aumentar o volume de analises a serem realizadas (Notas L e M - secção 3.2.4 - [65, 64]). Os resultados apresentados mostram que 47% das sinalizações realizadas se confirmaram ou carecem de aprofundamento nas investigações. Os 53% não confirmados podem ser reduzidos em função de um maior conhecimento da agora existente base histórica dos perfis comportamentais dos clientes. Estes números consideram a utilização ideal da solução apresentada, sendo que a instituição financeira conta com o recurso da Margem Adicional de Risco (MAR) que permite ajustar seu chamado apetite ao risco.

2. *Solução baseada em agentes, num processo de cooperação de análise, para tomada de decisão e mitigação do risco sistémico identificado no processo de ABC.*

   Considerando que uma mitigação eficiente do risco sistémico identificado neste trabalho é de difícil implementação, pois envolveria a atuação de órgãos reguladores ou de ações de governo, resta às instituições financeiras buscarem a melhoria contínua de seus processos e sistemas. Neste sentido, a solução implementada no âmbito do ABC, baseada em agentes inteligentes, resultante da aplicação de metodologia *Agent-Oriented Software Engineering (AOSE)*, suportada pela ferramenta *Prometheus Design Tool (PDT)*, cumpre o papel de ação mitigadora local. Os agentes implementados cumprem o papel de Analistas de primeira linha e estão aptos para decidir sobre o que passa ou não para os Analistas humanos. É uma decisão da instituição, por exemplo, definir que clientes selecionados que tenham baixo risco, não se enquadrem nas regras de reclassificação de risco e tenham baixo Índice de

Suspeição, sejam enquadrados como não suspeitos e retirados da relação a ser passada para os Analistas humanos. Esta flexibilidade é possível devido o ambiente de atuação, a arquitetura, forma de interação e ferramentas adicionais (Índice de Suspeição (IS), Mineração de Dados (MD), Margem Adicional de Risco) utilizadas na implementação do Sistemas Multi-Agente (SMA).

3. *Metodologia de utilização de técnicas de MD aplicável a grande volume de dados visando agrupamento de clientes, induzido sob uma perspetiva de grupos de risco.*

   A pesquisa sobre trabalhos relacionados foi evidenciada a opção de algumas soluções em basear-se, quase exclusivamente, em informações cadastrais. Parte dessas informações seguem exigências legais e uma parte significativa atendem uma boa prática internacional denominada *Know Your Customer (KYC)* (Notas D, P e Q - secções 3.1.4, 3.2.5 e 3.2.5 - [74, 69, 46]), ocorre que a própria banca reconhece a dificuldade em manter estas informações atualizadas. Torna-se, portanto, fundamental a busca por variáveis alternativas que permitam classificar ou agrupar os clientes. O Capítulo 4 apresenta uma sequência de etapas com descrição e volumetria dos dados utilizados, permitindo a utilização das técnicas escolhidas, seleção do melhor resultado e agrupamento dos clientes. Considerando que os atributos definidos são vocacionados para uma análise de risco, o resultado permitiu a formação de grupos de clientes com diferentes níveis de risco e sobre estes grupos foi possível montar a estratégia de tratamento para os problemas identificados. A sistematização das etapas citadas permitiu a criação de um algoritmo geral para o processo de aprendizagem (descrito em 4.3.3), incorporando os algoritmos envolvidos na solução e a escolha automática do melhor resultado.

4. *Modelo de integração entre a metodologia de MD e a arquitetura de sistema baseado em agentes Belief-Desire-Intention (BDI)*

   O modelo criado para o sistema apresenta os seguintes elementos básicos: a) um conjunto de regras obtidas como saída do algoritmo geral para o processo de aprendizagem; b) um conjunto menor de regras criadas com base na experiência dos Analistas de ABC da instituição financiadora da pesquisa; c) uma base histórica com o perfil comportamental dos clientes; e d) o perfil do cliente em análise. Este conhecimento foi codificado na forma de crenças para o modelo BDI, conforme descrito na secção 5.4. Conhecimento sobre o ambiente utilizado, valores que auxiliam a tomada de decisão sobre manuseio dos perfis, IS, MAR etc., foram codificados como Crenças Globais, permitindo a integração entre a MD e a arquitetura BDI.

5. *Resultados obtidos*

   Os resultados obtidos mostram a viabilidade do uso sistemático e estabelecem uma nova frente de combate ao crime de BC. A qualidade dos resultados foi atestada pela verificação realizada pelos analistas de ABC nos perfis suspeitas sinalizadas, nos quais vale destacar que *todos os casos classificados pelo sistema como suspeitos*

de "alto risco" não foram sinalizados anteriormente pelos sistemas em uso; dos 46 suspeitos totalmente confirmados, 35 nunca haviam sido relatados anteriormente por nenhum outro sistema em execução.

Além disso, calculando a precisão e as métricas F1-Score e *Matthews Correlation Coefficient (MCC)*, os resultados são claramente melhores no novo sistema. Considerando o centro dos intervalos analisados, conforme apresentado na Tabela 6.8 (secção 6.7), a melhora obtida foi de 19 pontos percentuais na métrica F1-score e 18 pontos na métrica MCC.

7.3 Publicações resultantes deste trabalho

O desenvolvimento e resultados obtidos nesta pesquisa possibilitaram a publicação em conferencias e artigos em jornais, uma curta descrição é apresentada a seguir:

- Alexandre, C. and J. Balsa 2015a. A Multiagent Based Approach to Money Laundering Detection and Prevention. In Proceedings of the International Conference on Agents and Artificial Intelligence, S. Loiseau, J. Filipe, B. Duval, and H. J. van den Herik, eds., volume 1, Pp. 230–235, Lisbon. SciTePress.

  Este artigo apresentou a proposta a ser desenvolvida na pesquisa, indicando a problemática envolvida tanto quanto a volumetria como também a necessidade de um maior apoio aos Analistas de ABC. O artigo indica como solução a utilização de um sistema de agentes inteligentes e descreve alguns dos agentes que integraram o sistema. Número de citações: 20[1]

- Alexandre, C. and J. Balsa 2015b. Client Profiling for an Anti-Money Laundering System. ArXiv e-prints. http://adsabs.harvard.edu/abs/2015arXiv151000878A.

  Este documento foi disponibilizado contendo a descrição detalhada do processo de mineração de dados para montagem de perfis de clientes bancários, visando apoiar o processo de deteção de operações suspeitas de BC. Contém detalhes dos experimentos realizados com os dados reais de uma instituição financeira, da forma de agrupamento dos clientes em *clusters* e da geração de um conjunto de regras de classificação, sinalizando como essas regras serão incorporadas na base de conhecimento dos agentes inteligentes responsáveis pela sinalização de transações suspeitas. Número de citações: 23[1]

- Alexandre, C. and J. Balsa 2016. Integrating client profiling in an anti-money laundering multi-agent based system. In New Advances in Information Systems and Technologies, Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, and M. Mendonça Teixeira, eds., Pp. 931–941, Cham. Springer International Publishing.

[1] Indexado por Google Académico Citações - posição verificada em 14 de agosto de 2022.

Este artigo resume o documento disponibilizado anteriormente e apresenta mais detalhes sobre os agentes definidos para o sistema multiagente e como eles utilizam os perfis criados e as regras geradas. Descreve a solução encontrada para informações que ofereciam complexidade ao processo de mineração de dados, quer seja pela sua volatilidade (débito e crédito em conta) quer seja pela amplitude dos valores envolvidos (de décimos a mil milhões da moeda envolvida). Número de citações: 22[1]

- Alexandre, C. and J. Balsa 2017. Um sistema multiagente no combate ao branqueamento de capitais. RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, 12(25):1 – 17.

  O algoritmo definido para automatização do processo de tratamento dos dados (seleção de transações relevantes), escolha dos melhores *clusters* e geração de regras foi apresentado neste artigo. Nele também foi apresentada a abordagem relacionada ao risco com a definição da Margem Adicional de Risco (MAR) que permite à instituição financeira ser mais ou menos conservadora em relação aos limites que definem a suspeição do cliente quanto a indícios de BC. Número de citações: 6[1]

- Alexandre, C. and J. Balsa 2018. A Multi-Agent System Based Approach to Fight-Financial Fraud: An Application to MoneyLaundering. Preprints.

  As técnicas utilizadas tanto no desenvolvimento do sistema quanto na sua implementação estão mais detalhadas neste documento. Os primeiros resultados obtidos com a implementação do sistema foram discutidos, demonstrando uma expressiva melhora na identificação de clientes suspeitos, inclusive indicando clientes não identificados por sistemas utilizados na instituição financeira. Número de citações: 4[1]

- Alexandre, C. and J. Balsa 2023. Incorporating Machine Learning and a Risk-based Strategy in an Anti-Money Laundering Multiagent System. Expert Systems with Applications, volume 217, Pp. 119500.

  Este artigo apresenta um resumo de toda a pesquisa realizada. Comenta o estudo realizado, o desenvolvimento e implementação do sistema, as técnicas e soluções adotadas. Apresenta a última forma do algoritmo de automatização do processo de tratamento dos dados, escolha dos melhores *clusters* e geração de regras, formaliza a MAR. Também define o Índice de Suspeição (IS) e discute o resultado final obtido. A superioridade dos resultados obtidos, quando comparados às soluções em uso na instituição financeira, foram demonstradas não só em valores absolutos mas pela utilização da métrica F1-score.

7.4 Trabalhos Futuros

O processo de Anti-Branqueamento de Capitais (ABC) continua carente de soluções que visem melhorar o processo de identificação e comunicação de clientes suspeitos do crime de Branqueamento de Capitais (BC). A seguir são sugeridas linhas de pesquisa que podem dar continuidade a este trabalho ou iniciarem novas linhas de pesquisa:

1. Buscar forma de codificar informação que permita maior nível de aprendizado dos agentes perante as decisões finais dos Analistas de ABC.

   A decisão do Analista vista de forma binária, comunica ou não comunica, casa haja ou não indício de Branqueamento de Capitais (BC), é a forma mais simples de codificação. Na realidade esta decisão concentra mais informações, tipo: quantidade de itens da regulação descumpridos; relevância e amplitude deste descumprimento; frequência com que o perfil foi comunicado; recomendação sobre a manutenção ou não da relação de negócio com o cliente envolvido por parte da instituição financeira. Todos estes itens envolvidos na decisão podem enriquecer a base de conhecimento e permitir um nível mais apurado de aprendizado e, consequentemente, de uma tomada de decisão mais automatizada ou, no mínimo, um auxílio mais qualificado para a tomada de decisão humana.

2. Incluir no modelo indicadores que permitam medir a eficácia das regras resultantes do aprendizado, diante da confirmação das sinalizações efetuadas.

   O conjunto de regras obtido é o elemento mais relevante para a seleção dos perfis suspeitos e posterior análise por parte dos Analistas de BC.Nem sempre a identificação e correção de falso positivo será suficiente para aprimorar a seleção, por isso, é grande valia que cada regra criada disponha de indicadores que permitam medir sua eficácia, possibilitando identificar, por exemplo, qual a qualidade da seleção realizada. Permitir medir, também, a efetividade da seleção, ou seja, dos perfis selecionados quantos foram comunicados e qual o grau de qualidade desta comunicação.

3. Modificar o modelo de regras utilizado, unificando-as ou transformando-as em cenários mais complexos de atuação dos branqueadores de capitais.

   O modelo de regras utilizado está fortemente ligado a itens da regulação vigente, contudo, além da conformidade legal por item, existem cenários que agrupam vários destes itens. Codificar estes cenários no modelo utilizado neste trabalho tornaria as regras muito complexas, principalmente se considerarmos a existência de cenários que vão além da legislação, ou seja, casos reais descobertos e divulgados e que ainda não estão tipificados. A codificação destes cenários seria utilizado em conjunto o modelo de regras utilizado neste trabalho.

4. Pesquisar um modelo eficiente de análise de resultados obtidos por soluções semelhantes a apresentada neste trabalho.

Testar, comparar e selecionar uma métrica adequada para problemas semelhantes ao apresentado neste trabalho seria uma grande contribuição. É possível até que o resultado deste estudo demonstre a necessidade de uma nova forma de avaliar resultados baseados em classificadores, que pode não conter as métricas escalares mais conhecidas. Dependendo da técnica ou conceito utilizado na solução do problema, qual a melhor métrica a ser utilizada para compara os resultados?

Apêndice A

Estrutura de Dados: Produto Contas-Correntes

Conta-corrente é o principal produto oferecido pela banca pública e privada, também conhecida como conta à ordem, pode ser mantida por pessoa singular ou jurídica e permite a realização de transações em espécie ou eletrónicas. Estas transações podem envolver depósitos, saques, transferências, pagamentos, tanto a débito quanto a crédito, nacionais ou internacionais. As principais tabelas do Sistema de Contas-Correntes estão abaixo descritas, a abreviatura colocada entre parênteses indica o nome utilizado no diagrama de dados (Figura A.1):

1. CADASTRO (ccad) – contém informações cadastrais dos clientes referentes ao sistema de contas-correntes. Cada cliente possui um único código a ele atribuído e que será associado com todas as contas correntes que este cliente possuir no banco. Não existe limite para a quantidade de contas que cada cliente pode abrir no banco, contudo, cada cliente tem apenas um único código de cliente a ele associado. Um código de cliente é gerado para cada conjunto [código da agência, número da conta-corrente, dígito verificador da conta] a menos que este conjunto pertença a um mesmo cliente e que já possui um código de cliente gerado, neste caso o conjunto é apenas associado ao código existente. Este código do cliente permite acesso a tabela com dados pessoais do cliente integrante do Sistema de Gerência do Cadastro;

2. MOVIMENTO (cmov) - armazena informações sobre todas as transações realizadas pelos clientes. É organizada pelo dia de realização da transação e pelo conjunto [código da agência, número da conta-corrente, dígito verificador da conta]. Não possui a informação de código do cliente. Possui um código referente a detalhes sobre a natureza da transação, chamado histórico, além dos valores envolvidos e o saldo da conta após a conclusão da transação;

3. HISTÓRICO (hist) - detém a descrição textual dos códigos que detalham a natureza

das transações realizada. Por este código é possível identificar se a transações foi de crédito ou débito, em dinheiro, por cheque, online, realizado numa agência ou por aplicativo móvel, envolveu transferência de recursos para outras agências do banco ou para outros bancos, etc.;

4. TIPO DE PESSOA (tpes) - define o tipo de "pessoa" do cliente. Neste caso "pessoa" refere-se a natureza contábil do cliente, caso ele seja pessoa física, jurídica, governo, empresa nos diversos portes, etc.;

5. TIPO DE CONTA (tcta) - determina o tipo de conta associada ao cliente. Podendo esta ser normal, detentora de cheque especial[1], conta salário, etc.;

6. TIPO CONTÁBIL (tctb) - relaciona a conta do cliente com um tipo contábil apropriado e que será utilizado pelo sistema de Contabilidade. Assume valores como "Instituição Financeira Cooperativa", "Governo Administração Direta Federal", etc.

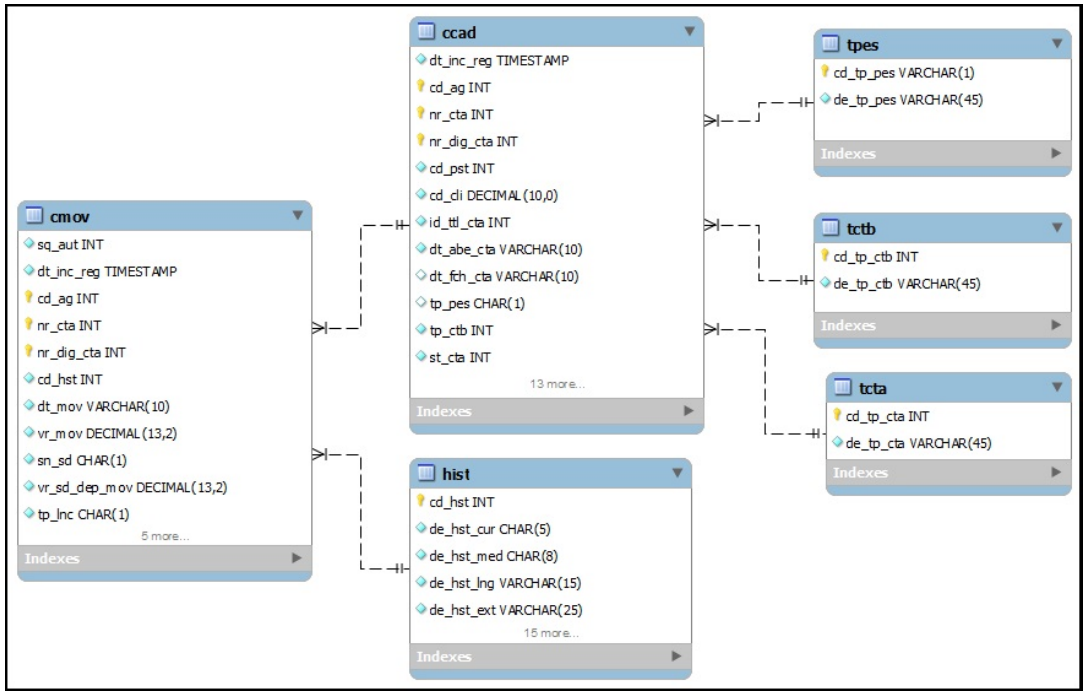


Figura A.1: Diagrama do Modelo de Dados

[1]Modalidade de empréstimo automático utilizado para na ocorrência de saques a descoberto. O valor disponível depende do limite estabelecido para cada cliente e é variável entre os bancos

Apêndice B

Diagramas da Modelagem do Sistema

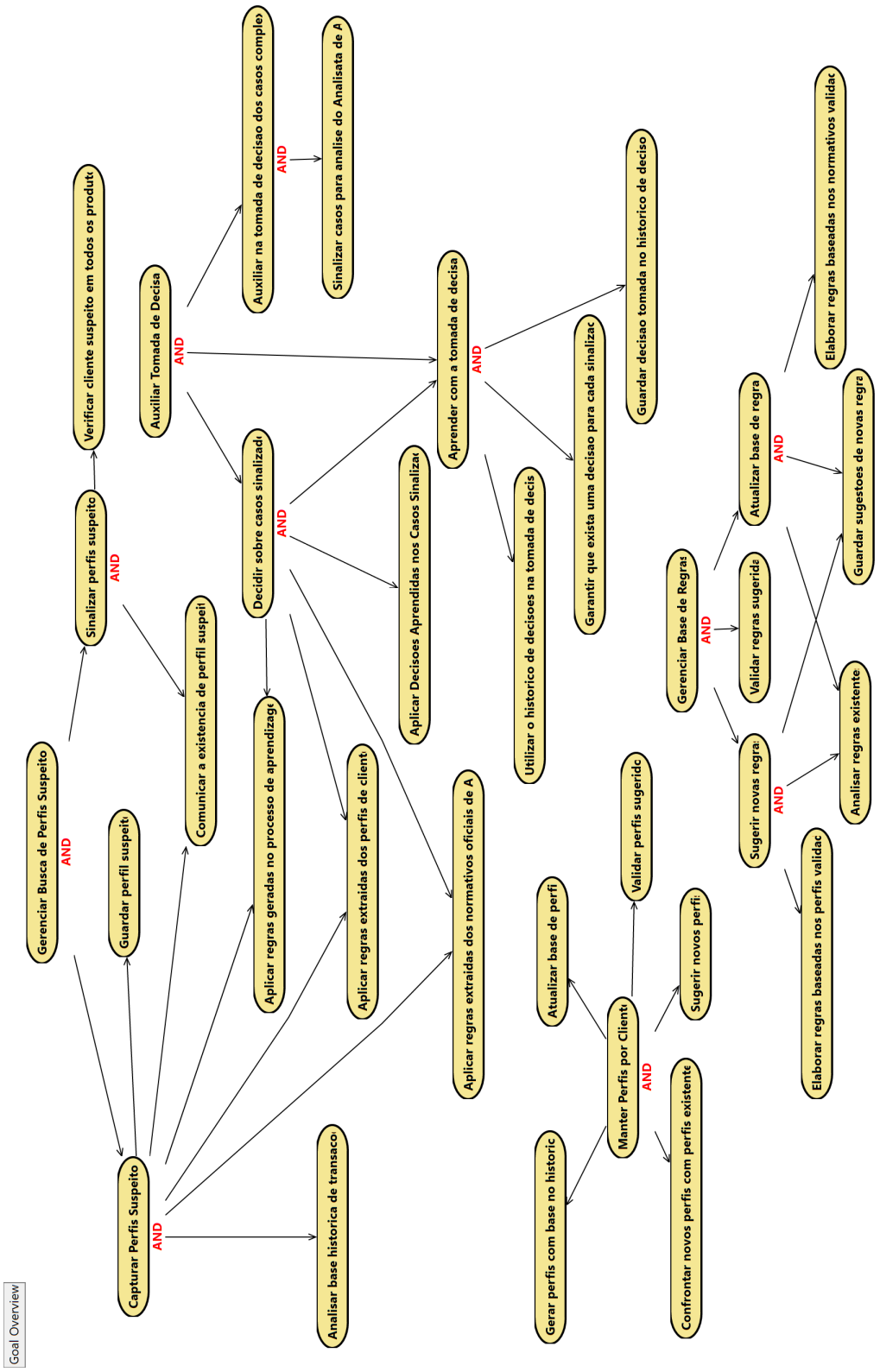


Figura B.1: Objetivos e Subobjetivos do Sistema

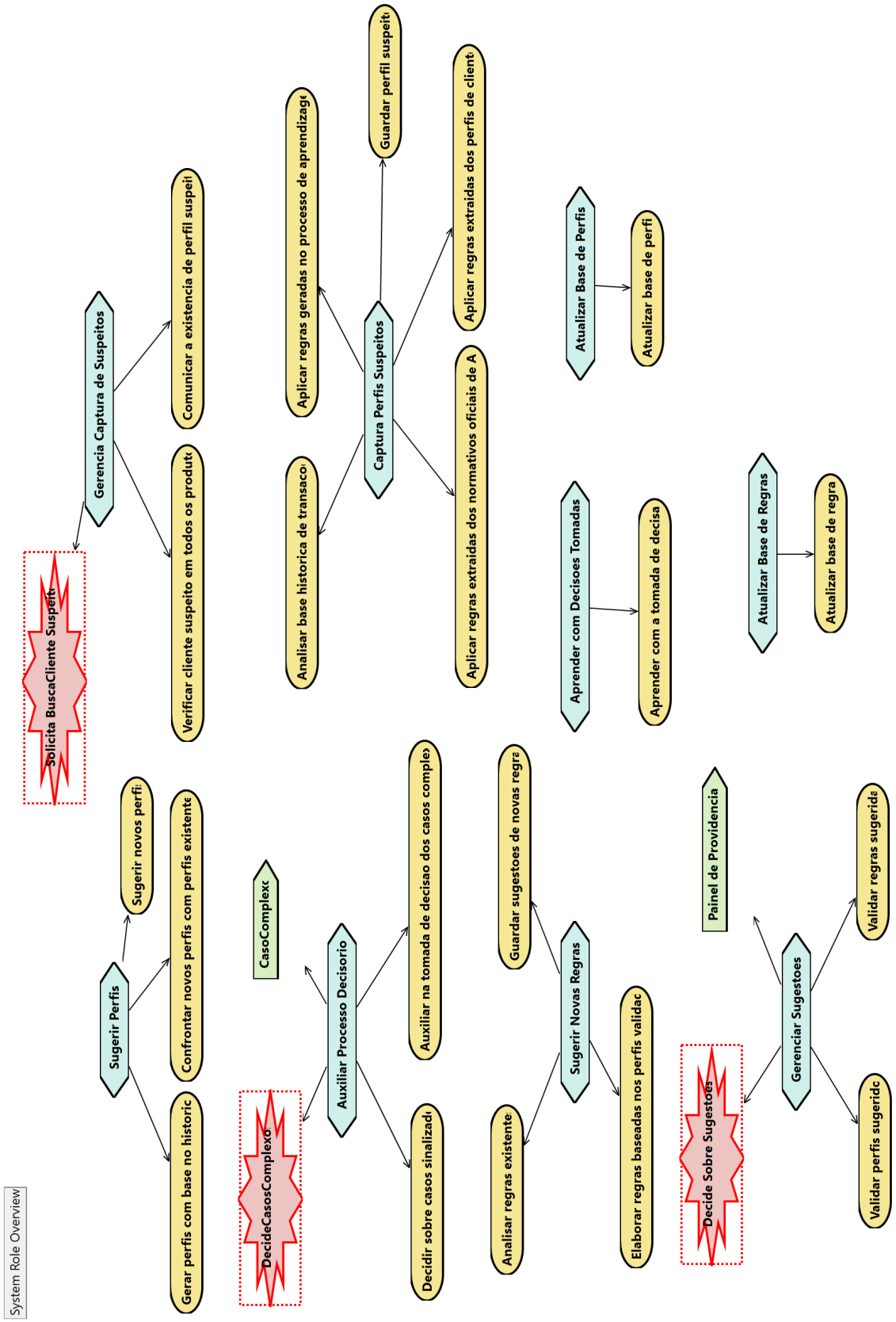


Figura B.2: Papeis Desempenhados no Sistema

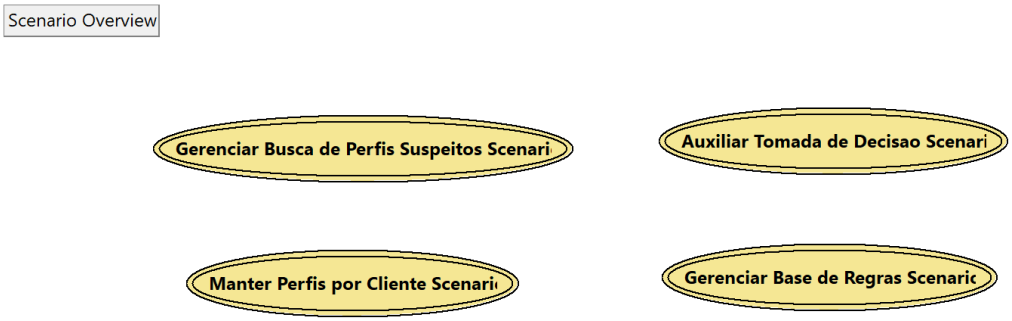


Figura B.3: Cenários do Sistema

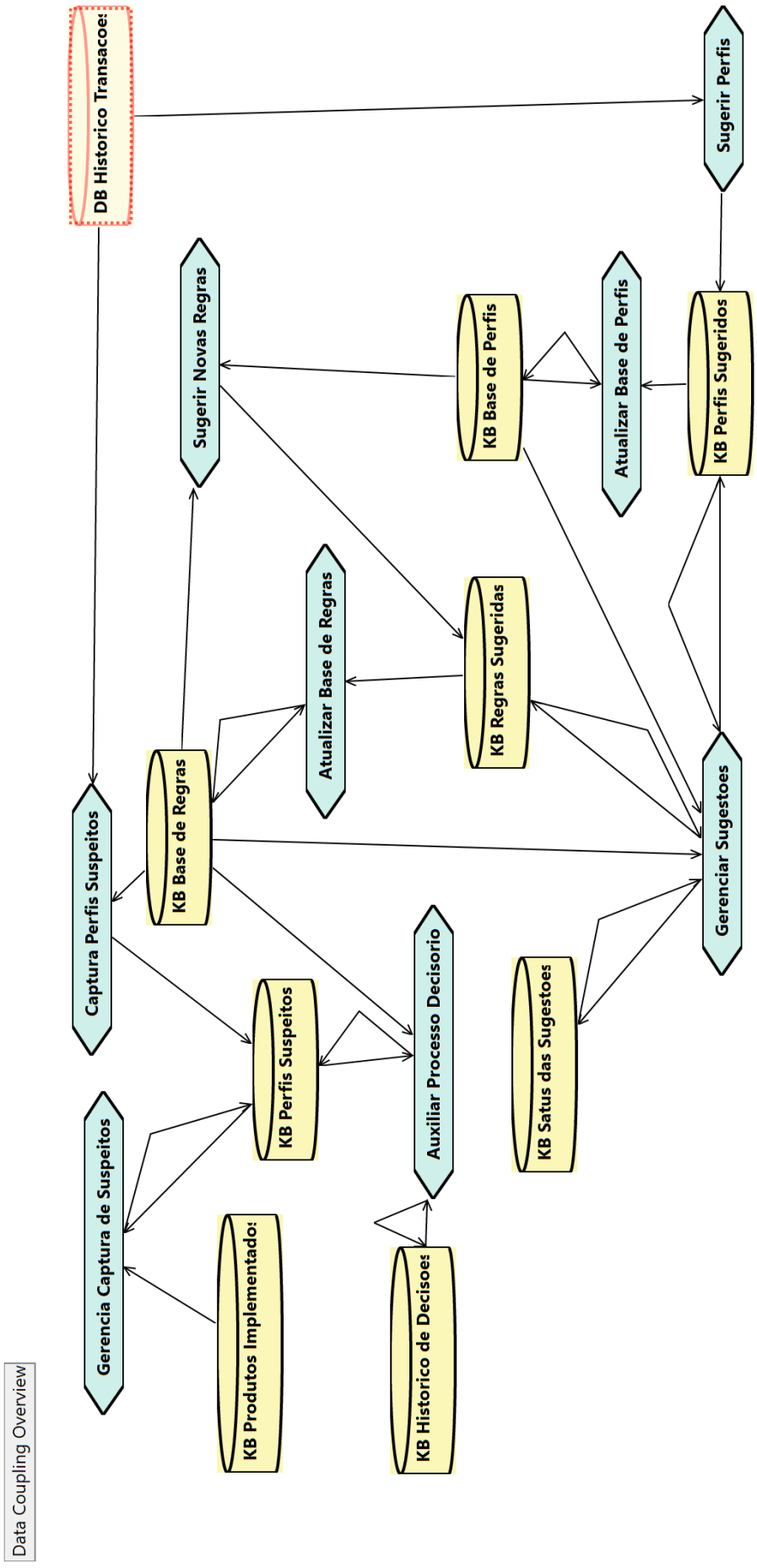


Figura B.4: Modelo de Dados do Sistema

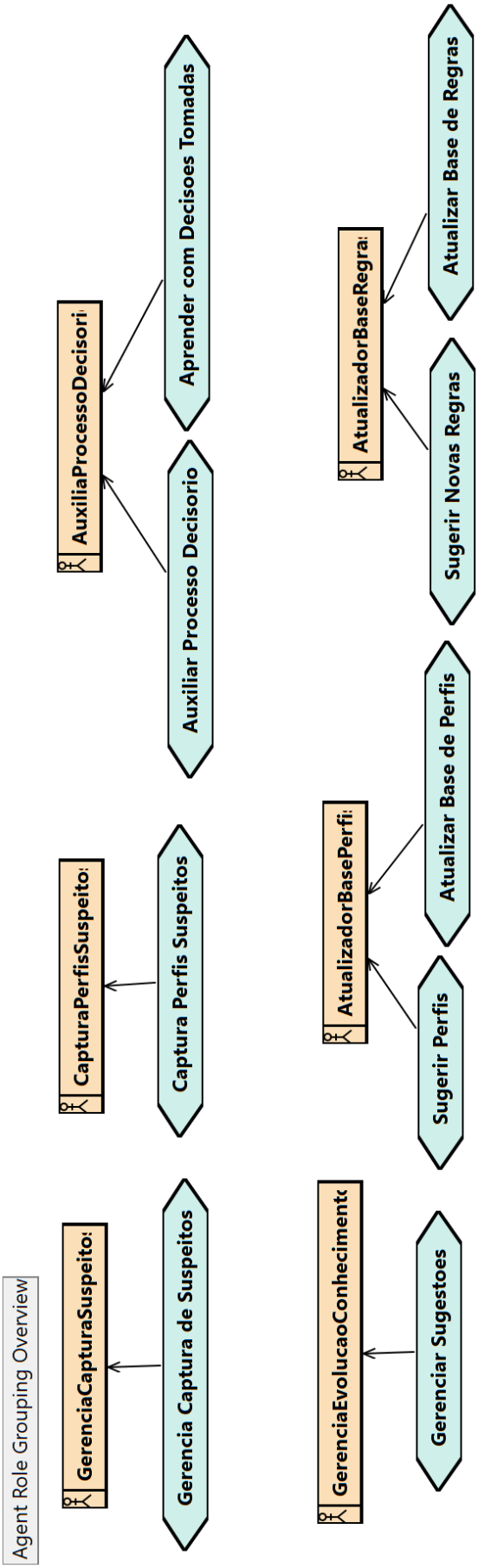


Figura B.5: Agrupamento de Papeis Desempenhados

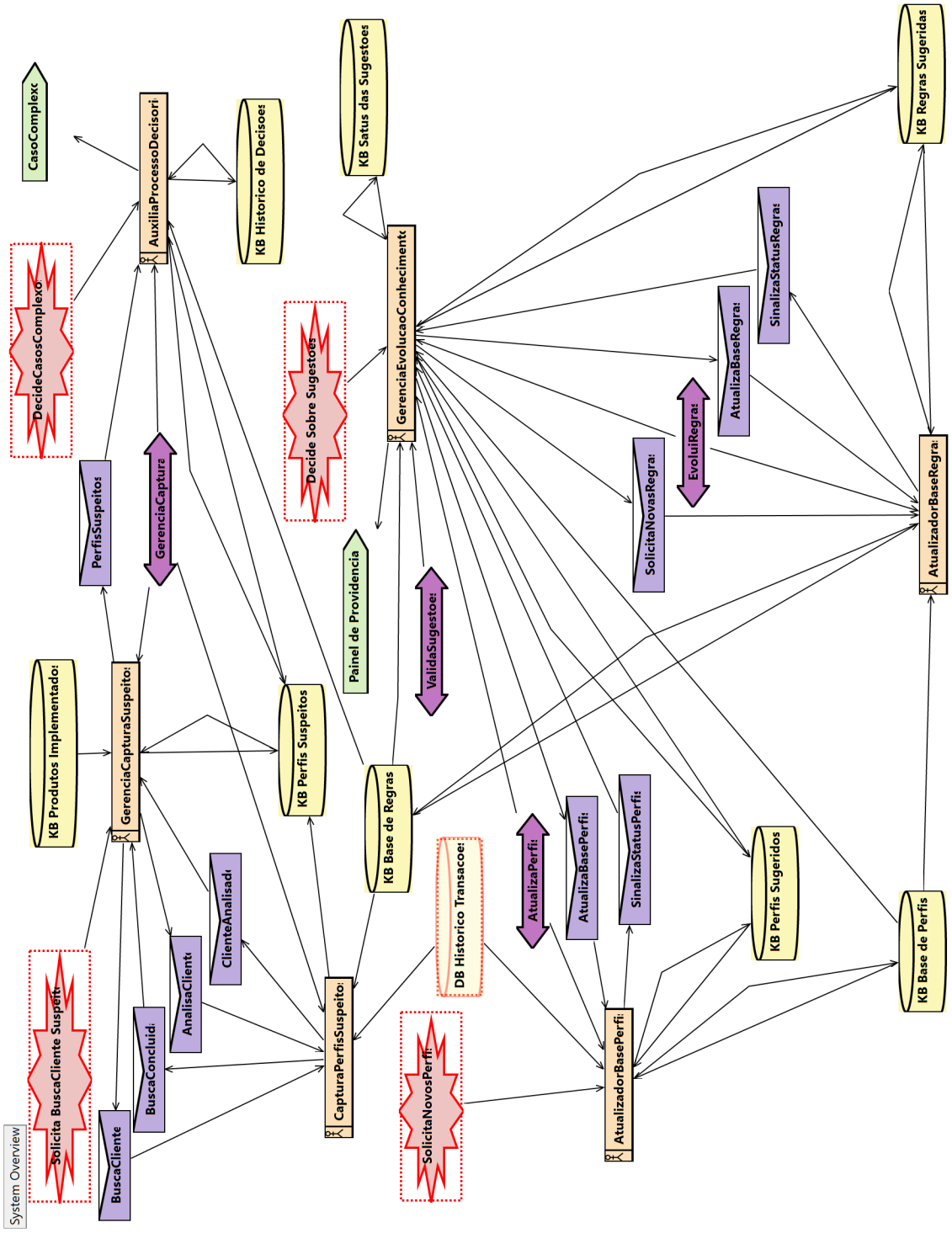


Figura B.6: Visão Geral do Sistema

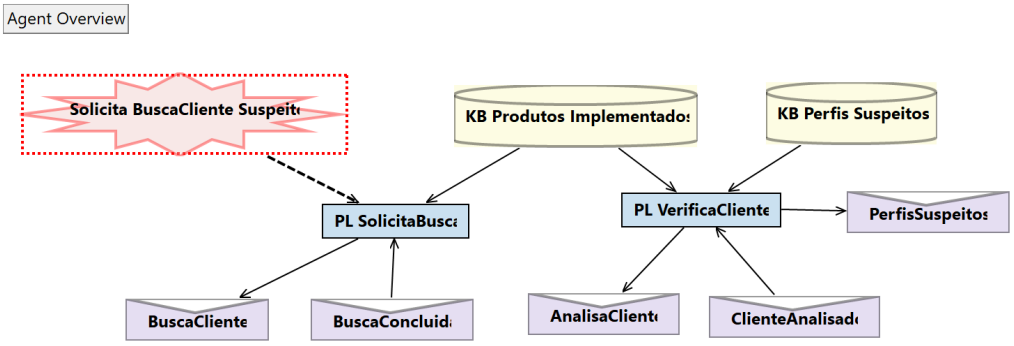


Figura B.7: Agente Gerenciador da Captura de Suspeitos

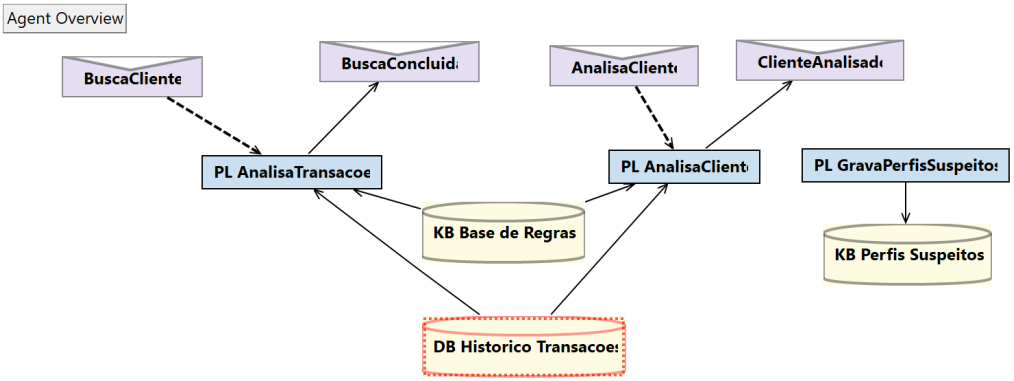


Figura B.8: Agente Capturador de Perfis Suspeitos

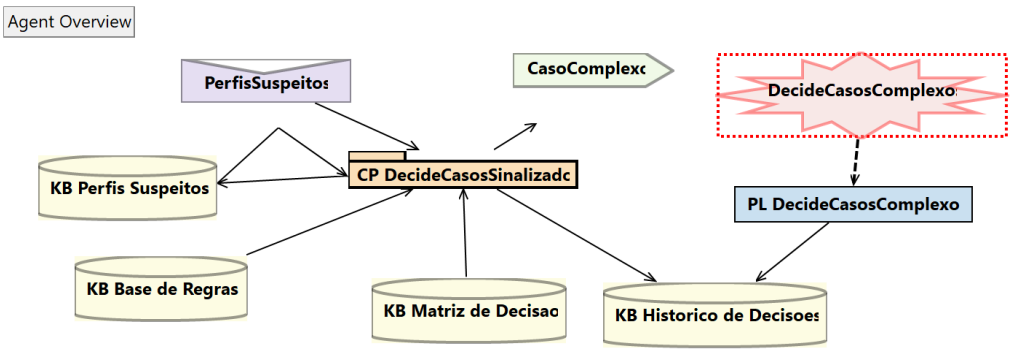


Figura B.9: Agente Auxilia Processo Decisório

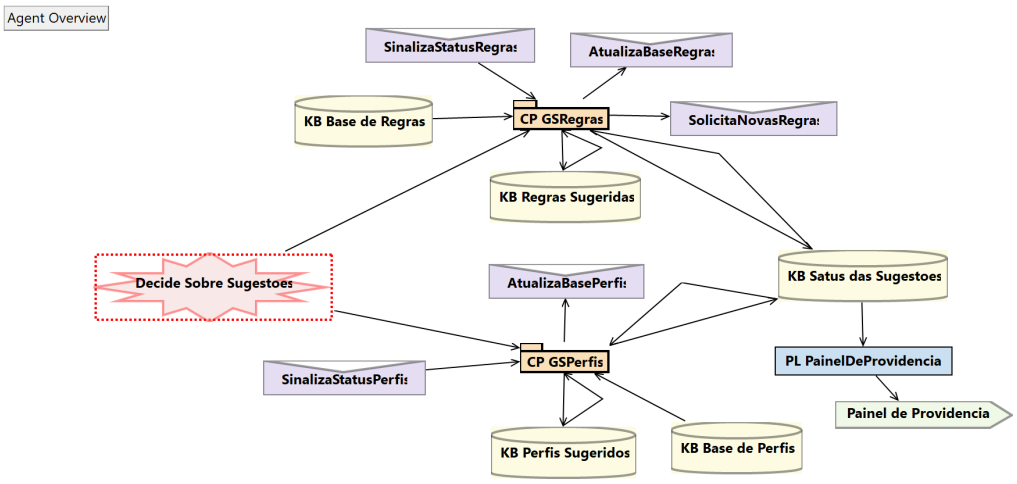


Figura B.10: Agente Gerenciador da Evolução do Conhecimento

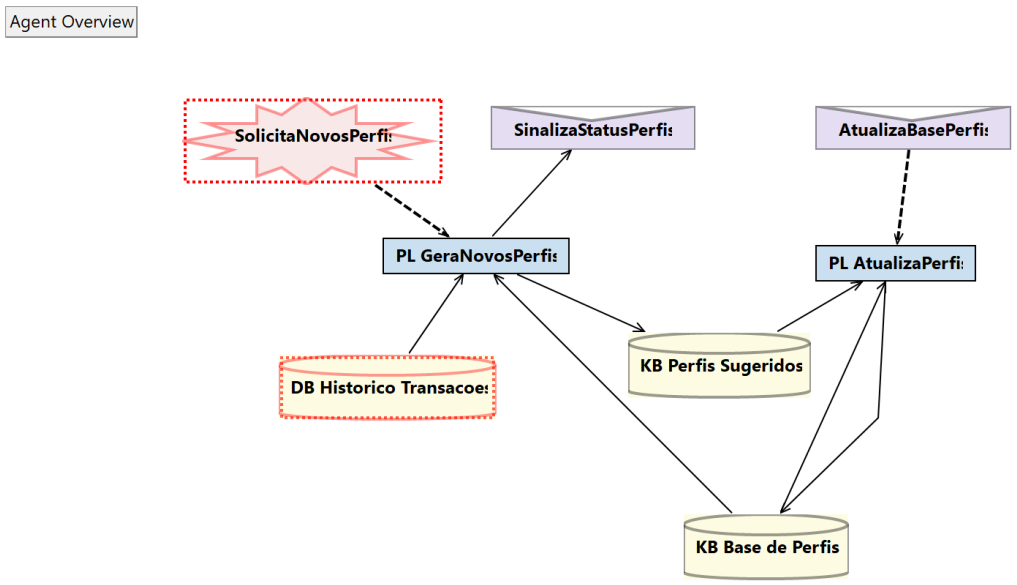


Figura B.11: Agente Atualizador da Base de Perfis

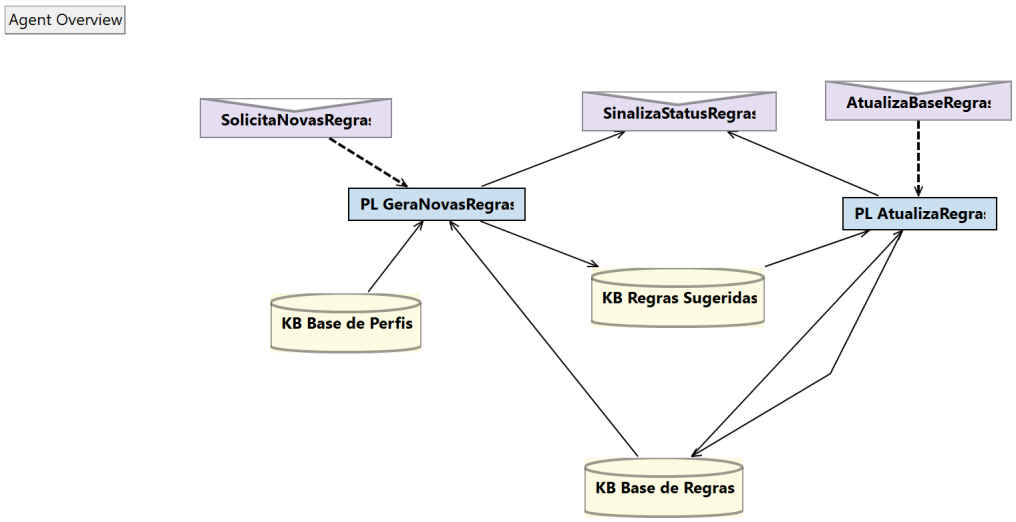


Figura B.12: Agente Atualizador da Base de Regras

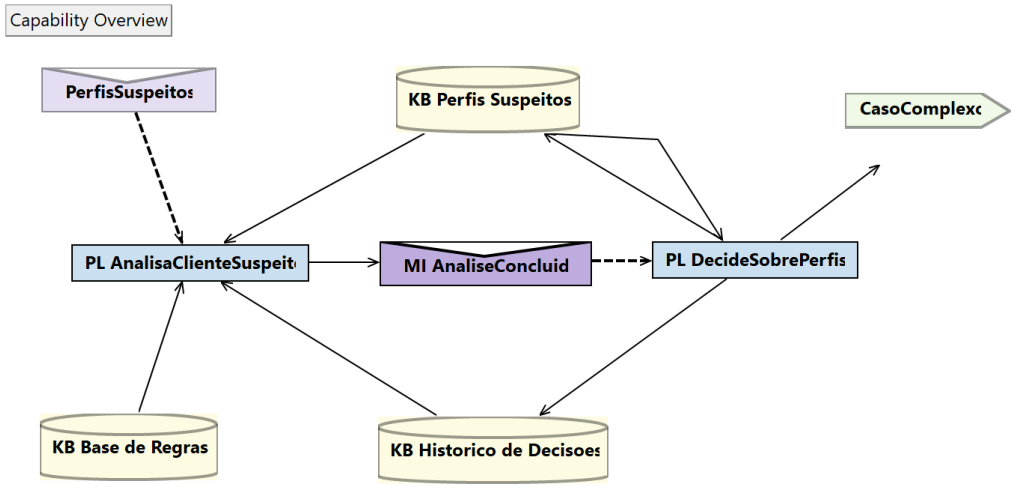


Figura B.13: Competência Decide Casos Sinalizados

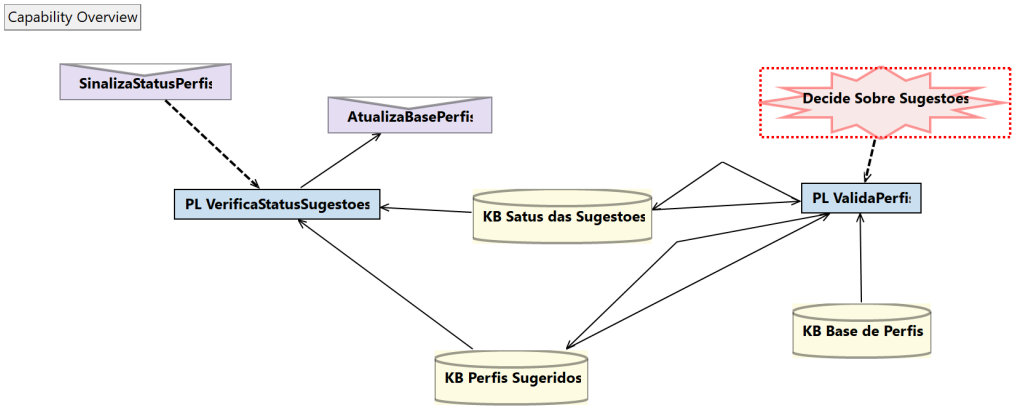


Figura B.14: Competência Gerencia Sugestão de Perfis

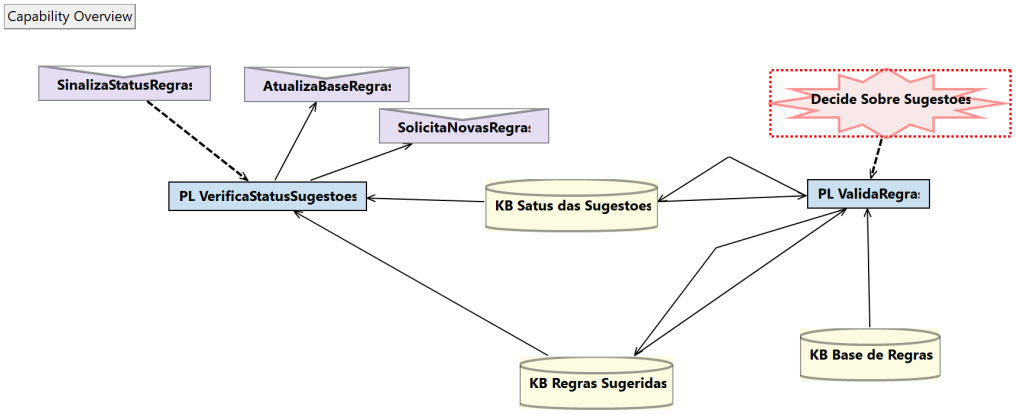


Figura B.15: Competência Gerencia Sugestão de Regras

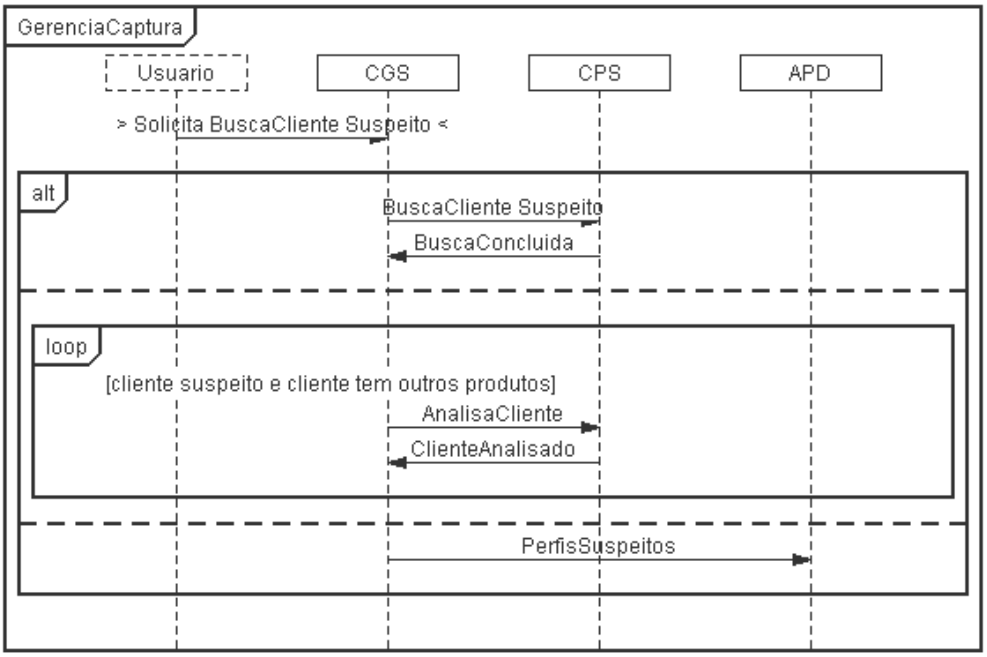


Figura B.16: Diagrama AUML Gerencia Captura de Suspeitos

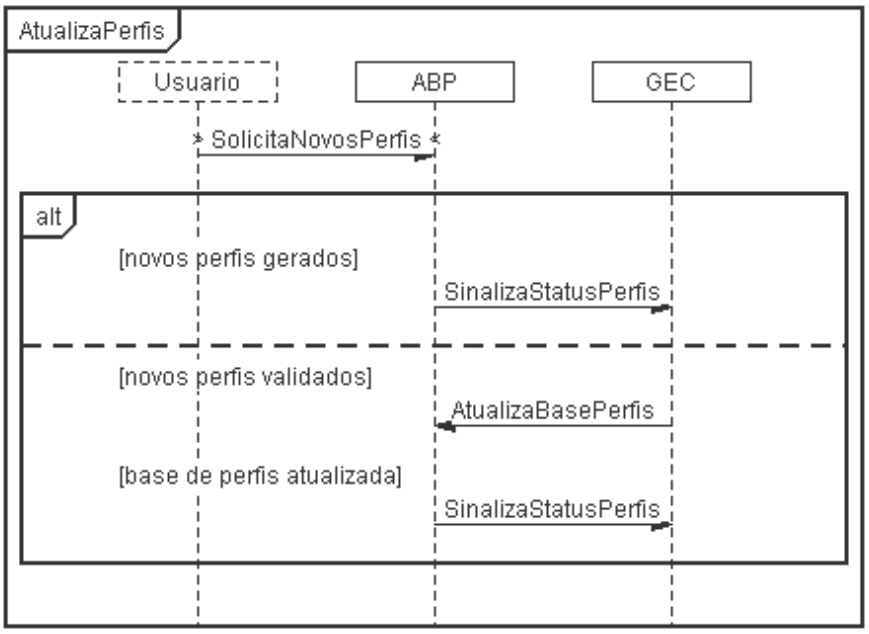


Figura B.17: Diagrama AUML Atualiza Perfis

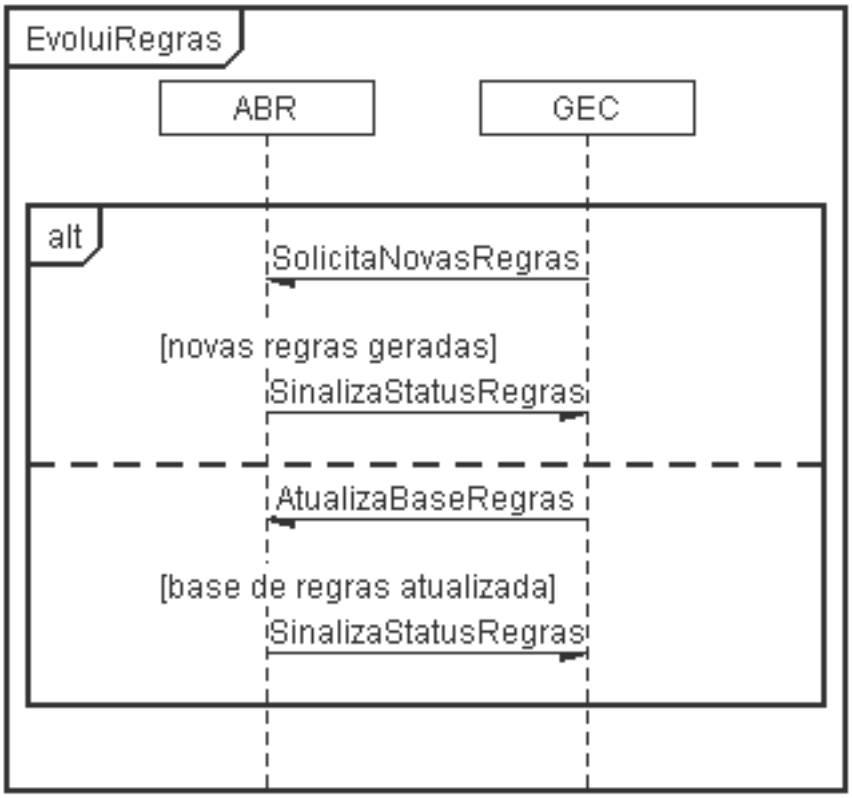


Figura B.18: Diagrama AUML Atualiza Perfis

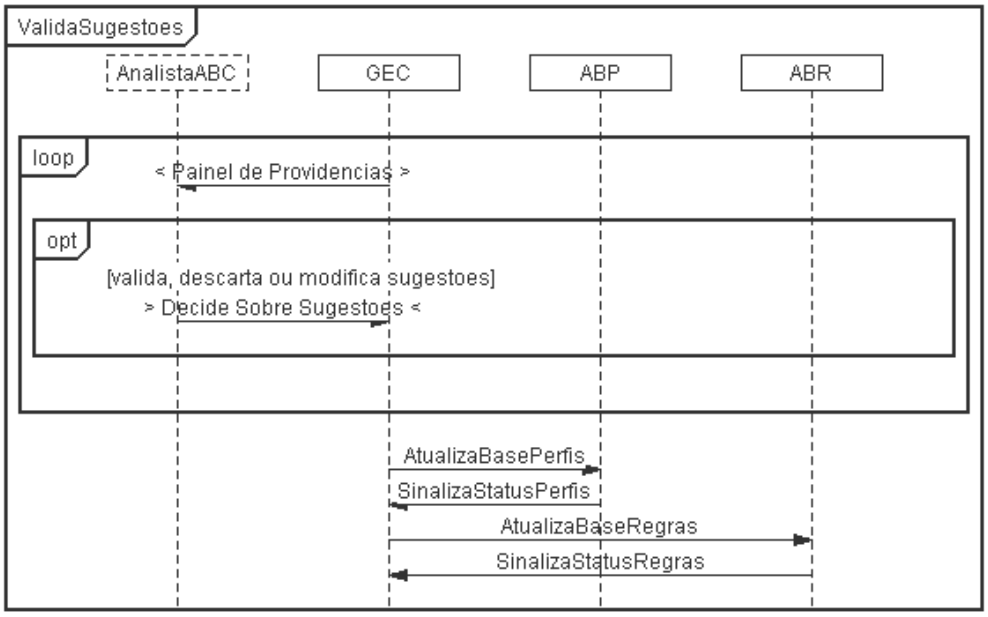


Figura B.19: Diagrama AUML Valida Sugestões

Apêndice C

Código-Fonte das Rotinas Mais Relevantes

```
/**
--Projeto:  Jano (Um MAS no Combate ao Branqueamento de Capitais)
--Arquivo Include
    crencasGlobais.asl - este arquivo contém informações sobre o ambiente
    de funcionamento do sistema e parâmetros de configuração que serão
    transformados em crenças para os diversos agentes.
--Localização
    Endereço "src/agt/inc/"
*/
// ************************************************************************
// Ambiente computacional
//
// caminho para abrigar os arquivos e relatórios de saída
dirArquivosSaida("C:/Users/.../JanoOutfiles/").

// quantidade de perfis por arquivo depende da quantidade de memória disponível
qtdePerfisPorAgente(40000).

// ************************************************************************
// Crenças específicas do produto Contas Correntes
//
// códigos que definem o cliente como Pessoa Física
tipoPessoa(Cod, pf)
    :- .member(Cod,["A", "B", "C", "D", "E", "F", "G", "H","N", "R", "S",
            "T", "V", "W", "<"]).
// códigos que definem o cliente como Outro Tipo de pessoa
tipoPessoa(Cod, ot)
    :- .member(Cod, ["I", "J", "K", "L", "M", "O", "P", "Q", "U", "X","Y",
            "Z", "1", "2", "3", "4", "5", "6", "7", "8", "9"]).
tipoPessoa(Cod, erro) :- true.
// nome dos perfis encontrados no processo de data mining
nomePerfis(["Baixa utilização","Padrão","Alerta","Risco","Alto Risco",
        "ErroClasse"]).
// perfis encontrados para os tipos de pessoa, conforme o nível de risco
defineClasse(pf, 0, cluster4, normal).  defineClasse(ot, 0, cluster4, normal).
defineClasse(pf, 1, cluster2, normal).  defineClasse(ot, 1, cluster1, normal).
defineClasse(pf, 2, cluster1, risco1).  defineClasse(ot, 2, na).//não agrupado
defineClasse(pf, 3, cluster5, risco2).  defineClasse(ot, 3, cluster3, risco2).
defineClasse(pf, 4, cluster3, risco3).  defineClasse(ot, 4, cluster2, risco3).
defineClasse(pf, 2, nclassif, risco1).  defineClasse(ot, 2, nclassif, risco1).

// ************************************************************************
// Crenças que afetam as regras de produção
//
// Margem de Risco(\%), que modifica o limite padrão do perfil permitindo ao
// utilizador flexibilizar sem modificar as regras de produção
pctMargemRisco(0).

// ************************************************************************
// Crenças que afetam a análise e apresentação dos resultados
//
// pesos dos 11 atributos utilizados no cálculo do Índice de Suspeição
pesosIndiceSuspeicao([1,1,1,1,1,2,1,1,1,2,2]).
```

Código C.1: Crenças Globais

```
// controle para analisar todos os perfis selecionados
for ( .range(NumPerfil, 1, QtdePerfis) ) {
    -+numPerfil(NumPerfil);
    ArqAtual = ((NumPerfil - 1) div QtdePerfisPorAgente);   //limite para
    ArquivoAtual = (ArqAtual * QtdePerfisPorAgente) + 1;    //criação de
    if ( agenteCPS(NumPerfil, NomeAgente) ) {           //novo agente
        !atualizaArquivoSuspeitos;
        if ( agenteAnterior(NomeAgtAnt) ) {
            .kill_agent(NomeAgtAnt);            // elimina agente anterior
            retiraAgenteGCSgui(NomeAgtAnt); // atualiza interface
        }
        .create_agent(NomeAgente, "recuperadorPerfisAnalise.asl",
                    [beliefBaseClass("jason.bb.TextPersistentBB")]);
        -+agenteAnterior(NomeAgente);
    } else {
        ?agenteCPS(ArquivoAtual, NomeAgente);
    }
    // solicita perfil ao agente criado e adiciona crenças com os dados
    .send(NomeAgente, askOne, NumPerfil, PerfilCliente);
    !obtemDadosPerfil(PerfilCliente);
    // aplica os conjuntos de regras ao perfil em análise
    ?codTipoPessoa(Codigo);
    ?classe(ClasseOriginal);
    !redefineClasse(NumPerfil, ClasseDM);
    !aspectosLegais(NumPerfil, ClasseDM);
    !limitesPerfil(NumPerfil, ClasseDM);
    !memorizaSeSuspeito(NumPerfil);
    // totaliza classe original/reclassificação por tipo de pessoa
    !contaClasse(Codigo, ClasseOriginal, ClasseDM);
}
```

Código C.2: Processo Principal

```
// aplica regras obtidas no DM visando correção de erros da matrix de confusão
+!redefineClasse(PerfilCliente, ClasseRedefinida)
    :   codTipoPessoa(Cod) & tipoPessoa(Cod, Tipo) & Tipo == pf
    <-  ?regraPF(PerfilCliente, ClasseRedefinida, IdRegra);
        !guardaIdRegra(IdRegra).
+!redefineClasse(PerfilCliente, ClasseRedefinida)
    :   codTipoPessoa(Cod) & tipoPessoa(Cod, Tipo) & Tipo == ot
    <-  ?regraOT(PerfilCliente, ClasseRedefinida, IdRegra);
        !guardaIdRegra(IdRegra).
+!redefineClasse(PerfilCliente, ClasseRedefinida)
    <- ?classe(ClasseRedefinida);
        !guardaIdRegra("DMxxErro01").
// aplica regras extraídas dos normativos dos órgãos de controle
+!aspectosLegais(NumPerfil, Classe)
    :   codTipoPessoa(Cod) & tipoPessoa(Cod, Tipo) &
        defineClasse(Tipo, _, Classe, Status)
    <-  ?regraBC(ResultAnalise, Status, IdRegra);
        !guardaIdRegra(IdRegra).

-!aspectosLegais(NumPerfil, Classe) : true. // perfil com erro
```

Código C.3: Reclassificação dos Perfis

```
1  // Encerra se o cliente já pertencer a grupo de risco
2  regraPF(Pcli, Classe, IdRegra) :- classe(ClasseAtual) & idxStatusClasse(pf, Idx,
       ClasseAtual, Status) &
3      .substring("risco",Status) & Classe = ClasseAtual.
4  regraPF(Pcli, Classe, IdRegra) :- classe(ClasseAtual) & idxStatusClasse(pf, Idx,
       ClasseAtual, Status) &
5      Status == erro & Classe = ClasseAtual & IdRegra = "DMPFErro01".
6
7  // Cliente Grupo Alerta - Classe 1
8  regraPF(Pcli, cluster1, "DMPF2016001") :- pctDebito(PctDebito) & pctTED(PctTED)
       &
9      PctDebito > 76.69 & PctTED > 74.24.
10 regraPF(Pcli, cluster1, "DMPF2016002") :- pctDebito(PctDebito) & pctTED(PctTED)
       & qtdeServico(QtdeServico) &
11     PctDebito > 97.35 & PctTED > 73.48 & PctTED <= 74.24 & QtdeServico <= 48.
12 .
13 .
14 .
15 // Cliente Grupo Alto Risco - Classe 3
16 regraPF(Pcli, cluster3, "DMPF2016016") :- pctDebito(PctDebito) & pctTED(PctTED)
       & qtdeServico(QtdeServico) &
17     idadeConta(IdadeConta) &
18     PctDebito > 51.92 & PctTED <= 25.25 & IdadeConta > 11 & QtdeServico > 38.
19 regraPF(Pcli, cluster3, "DMPF2016017") :- pctDebito(PctDebito) & pctTED(PctTED)
       & qtdeServico(QtdeServico) &
20     idadeConta(IdadeConta) &
21     PctDebito > 51.92 & PctTED <= 25.25 & IdadeConta <= 11 & QtdeServico > 64.
22 regraPF(Pcli, cluster3, "DMPF2016018") :- pctDebito(PctDebito) & pctTED(PctTED)
       & qtdeServico(QtdeServico) &
23     idadeConta(IdadeConta) &
24     PctDebito > 51.92 & PctTED <= 25.25 & IdadeConta > 16 & QtdeServico > 21 &
           QtdeServico <= 38.
25 regraPF(Pcli, cluster3, "DMPF2016019") :- pctDebito(PctDebito) & pctTED(PctTED)
       & qtdeServico(QtdeServico) &
26     idadeConta(IdadeConta) &
27     PctDebito > 100 & PctTED <= 25.25 & IdadeConta <= 11 & QtdeServico > 55 &
           QtdeServico <= 64.
28 regraPF(Pcli, cluster3, "DMPF2016020") :- pctDebito(PctDebito) & pctTED(PctTED)
       & qtdeServico(QtdeServico) &
29     idadeConta(IdadeConta) &
30     PctDebito > 100 & PctTED <= 25.25 & IdadeConta > 6 & IdadeConta <= 11 &
           QtdeServico > 42 & QtdeServico <= 55.
```

Código C.4: Exemplo de Tratamento de Regras

Referências Bibliográficas

[1] Alexandre, C. and J. Balsa

2015a. A Multiagent Based Approach to Money Laundering Detection and Prevention. In *Proceedings of the International Conference on Agents and Artificial Intelligence*, S. Loiseau, J. Filipe, B. Duval, and H. J. van den Herik, eds., volume 1, Pp. 230–235, Lisbon. SciTePress.

[2] Alexandre, C. and J. Balsa

2015b. Client Profiling for an Anti-Money Laundering System. *ArXiv e-prints*. `http://adsabs.harvard.edu/abs/2015arXiv151000878A`.

[3] Alexandre, C. and J. Balsa

2016. Integrating client profiling in an anti-money laundering multi-agent based system. In *New Advances in Information Systems and Technologies*, Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, and M. Mendonça Teixeira, eds., Pp. 931–941, Cham. Springer International Publishing.

[4] Alexandre, C. and J. Balsa

2017. Um sistema multiagente no combate ao branqueamento de capitais. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, 12(25):1 – 17.

[5] Alexandre, C. and J. Balsa

2018. A Multi-Agent System Based Approach to FightFinancial Fraud: An Application to MoneyLaundering. *Preprints*.

[6] Alexandre, C. R. and J. Balsa

2023. Incorporating machine learning and a risk-based strategy in an anti-money laundering multiagent system. *Expert Systems with Applications*, 217:119500.

[7] AlixPartners Team

2018. 2018 Global Anti-Money Laundering and Sanctions Compliance Survey. techreport, AlixPartners. Acessada em dd/mm/aaaa.

[8] Allan, R.

2010. Survey of Agent Based Modelling and Simulation Tools. techreport DL-TR-2010-007, Science and Technology Facilities Council.

[9] Arthur, D. and S. Vassilvitskii
2007. K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Pp. 1027–1035, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics.

[10] Baldoni, M., C. Baroglio, F. Capuzzimati, and R. Micalizio
2015. Programming with commitments and goals in jacamo+. In *Proceedings of the 2015 International Conference AAMAS*, Pp. 1705–1706, Richland, SC.

[11] Balke, T. and N. Gilbert
2014. How Do Agents Make Decisions? A Survey. *Journal of Artificial Societies and Social Simulation*, 17(4):1–13.

[12] Banco de Portugal
2009. Supervisão prudencial - branqueamento de capitais. `http://www.bportugal.pt/pt-PT/Supervisao/SupervisaoPrudencial/BranqueamentoCapitaisFinanciamentoTerrorismo/Paginas/branqueamentodecapitais.aspx`. Acessado em 23/11/2015.

[13] Banco de Portugal
2013. Políticas de inclusão e formação financeira. Departamento de Supervisão Comportamental. `https://clientebancario.bportugal.pt/sites/default/files/2017-10/PoliticasInclusaoFormacaoFinanceira2013.pdf`, Acessado em 15/12/2021.

[14] Banco de Portugal
2018. Bpstat estatísticas online. `http://www.bportugal.pt/EstatisticasWeb/(S(oggjsg452hpfoyalpcurgzuh))/Default.aspx`. Acessado em 15/12/2018.

[15] Barros, M. A.
1998. *Lavagem de Dinheiro. Implicações Penais, Precessuais e Administrativas*. São Paulo: Editora Oliveira Mendes.

[16] Basel Committee
1988. International convergence of capital measurement and capital standards - basle capital accord. Technical report, Bank for International Settlements.

[17] Basel Committee
2001. Customer due diligence for banks. Technical report, Working Group on Cross-border Banking - Basel Committee on Banking Supervision - Bank for International Settlements.

[18] Bawa, A. and V. K. Attri
2015. A study of tools used in implement agent oriented software engineering. *International Journal of innovative Research in Computer and Communication Engineering*, 3(5). `http://www.ijircce.com/upload/2015/may/40_A_Study.pdf`, Acessado em 15/09/2015.

[19] Bifet, A., R. Gavaldà, G. Holmes, and B. Pfahringer
2018. *Machine Learning for Data Streams with Practical Examples in MOA*. Cambridge, Massachusetts London, England Piscataqay, New Jersey: The MIT Press.

[20] Boissier, O., R. Bordini, J. Hübner, A. Ricci, and A. Santi
2013. Multi-agent oriented programming with jacamo. *Sci. Comput. Program.*, 78(6):747–761.

[21] Bolton, R. J. and D. J. Hand
2002. Statistical fraud detection: A review. *Statist. Sci.*, 17(3):235–255.

[22] Bordini, R. H. and J. Dix
2013. Programming multiagent systems. In *Multiagent Systems*, G. Weiss, ed., chapter 13, Pp. 587–639. London, England: The MIT-Press.

[23] Bordini, R. H., J. F. Hübner, and M. Wooldridge
2007. *Programming Multi-Agent Systems in AgentSpeak Using Jason (Wiley Series in Agent Technology)*. John Wiley & Sons.

[24] Bratman, M.
1987. *Intention, Plans, and Practical Reason*. Harvard University Press.

[25] Brooks, R. A.
1991. Intelligence without Representation. *Artificial Intelligence*, 47:139–159.

[26] Caliński, T. and J. Harabasz
1974. A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27.

[27] Camilo, C. O.
2009. Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas. techreport RT-INF 001-09, Instituto de informática - UFG, Goiás.

[28] Canas, V.
2004. *O Crime de Branqueamento: Regime de Prevenção e de Repressão*. Coimbra-PT: Ed. Almedina. 335p.

[29] Carvalho, F. J. C. d. et al.
2015. *Economia monetária e financeira: teoria e política*, 3a edition. Ed. Campus.

[30] Castellar, J. C.
2004. *Lavagem de Dinheiro - A Questão do Bem Jurídico*. Rio de Janeiro: Ed. Revan.

[31] Castillo-Rojas, W., F. Medina-Quispe, and J. Vega-Damke
2017. Esquema de Visualización para Modelos de Clústeres en Minería de Datos. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, Pp. 67 – 80.

[32] Chopra, A. K. and M. P. Singh
2013. *Agent Communication*, chapter 3, Pp. 101–141. The MIT Press.

[33] COAF
2005. *Lavagem de Dinheiro: Legislação Brasileira*. São Paulo: Febraban. 378p.

[34] COAF
2015. Cartilha - lavagem de dinheiro - um problema mundial. Cartilha, Conselho de Controle de Atividades Financeiras, Brasil. `http://www.coaf.fazenda.gov.br/menu/pld-ft/publicacoes/cartilha.pdf/view`, Acessado em 23/11/2015.

[35] COAF
2021. COAF em Números. Online. Acessado em jan/2021.

[36] Coelho, H.
2008. Teoria da Agência: Arquitetura e Cenografia.

[37] Cohen, W. W.
1995. Fast effective rule induction. In *In Proceedings of the Twelfth International Conference on Machine Learning*, Pp. 115–123, Lake Tahoe, CA. Morgan Kaufmann. `http://www.cs.utsa.edu/~bylander/cs6243/cohen95ripper.pdf`.

[38] Costa, E. and A. Simões
2008. *Inteligência Artificial - Fundamentos e Aplicações*, 3a edition. Lisboa, PT: FCA - Editora de Informática.

[39] da Silva Praça Gomes Pereira, I. C. C.
2004. *Sistema Multi-Agente para Apoio à Negociação em Mercados de Eletricidade*. phdthesis, Universidade de Trás-os-Montes e Alto Douro, Vila Real, PT.

[40] de Jesús Rocha-Salazar, J., M.-J. Segovia-Vargas, and M. del Mar Camacho-Miñano
2020. Money laundering and terrorism financing detection using neural networks and an abnormality indicator. *Expert Systems with Applications*, 169:114470.

[41] de Silva, L., L. Padgham, and S. Sardina
2019. HTN-like solutions for classical planning problems: An application to BDI agent systems. *Theoretical Computer Science*, 763:12–37.

[42] Decreto-Lei Portugal
2002. Decreto-lei nº 304/2002, de 13 de dezembro. Diário da República -

I Série-A. Assembleia da República, http://www.dgpj.mj.pt/sections/leis-da-justica/pdf-ult/decreto-lei-n-304-2002/.

[43] Decreto-Lei Portugal
2009. Decreto-lei nº 42/2009, de 12 de fevereiro. Diário da República - I Série. Assembleia da República, http://www.dgpj.mj.pt/sections/leis-da-justica/pdf-ult2/decreto-lei-42-2009.

[44] Demazeau, Y.
1995. From interactions to collective behaviour in agent-based systems. In *1st. European Conference on Cognitive Science. Saint-Malo*, Pp. 117–132. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.47.7968&rep=rep1&type=pdf, Acessado em 24/01/2014.

[45] Demetis, D. S.
2018. Fighting money laundering with technology: A case study of bank x in the UK. *Decision Support Systems*, 105:96–107.

[46] Demetis, D. S. and I. O. Angell
2007. The risk-based approach to AML: representation, paradox, and the 3rd directive. *Journal of Money Laundering Control*, 10(4):412–428.

[47] Dignum, V. and J. Padget
2013. *Multiagent Organizations*, chapter 2, Pp. 51–98. Cambridge, Massachusetts London, England: The MIT Press.

[48] Dow Jones Risk and Compliance Team and SWIFT Financial Crime Compliance Team
2017. Global Anti-Money Laundering Survey Results 2017. techreport, Dow Jones & Company Inc.

[49] Dreżewski, R., J. Sepielak, and W. Filipkowski
2012. System supporting money laundering detection. *Digital Investigation*, 9(1):8 – 21.

[50] Eurostat
2010. Money laundering in europe. Publications office of the european union, European Commission, Luxembourg. http://ec.europa.eu/eurostat/en/web/products-statistical-working-papers/-/KS-RA-10-003, Acessado em: 11/11/2015.

[51] Evertsz, R., J. Thangarajah, N. Yadav, and T. C. Ly
2015. Agent oriented modelling of tactical decision making. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems 2015 (AAMAS)*, R. H. Bordini, E. Elkind, G. Weiss, and P. Yolum, eds.,

Pp. 1051–1060. ACM. http://www.aamas2015.com/en/AAMAS_2015_USB/aamas/p1051.pdf, Acessado em 05/01/2016.

[52] Faceli, K., A. C. Lorena, J. Gama, and A. C. P. d. L. F. d. Carvalho
2022. *Inteligência artificial: uma abordagem de aprendizado de máquina*, 2 edition. Rio de Janeiro: LTC.

[53] FATF
2017. Anti-money laundering and counter-terrorist financing measures - Portugal. techreport, The Financial Action Task Force, Paris.

[54] Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth
1996. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37.

[55] FEBRABAN
2021. Pesquisa febraban de tecnologia bancária 2021. Technical report, São Paulo. https://cmsarquivos.febraban.org.br/Arquivos/documentos/PDF/pesquisa-febraban-relatorio.pdf, Acessado em dez/2021.

[56] Ferguson, R. W.
2003. Capital standards for banks: the evolving Basel Accord. *Federal Reserve Bulletin*, 89(9):395–405. https://www.federalreserve.gov/pubs/bulletin/2003/0903lead.pdf.

[57] Fernandez, A.
2019. Artificial Intelligence in Financial Services. *Banco de Espana Article*, (3/19).

[58] Firdaus, S. and M. A. Uddin
2015. A survey on clustering algorithms and complexity analysis. *International Journal of Computer Science Issues (IJCSI)*, 12(2):62.

[59] Fisher, M., R. Bordini, B. Hirsch, and P. Torroni
2007. Computational logics and agents - a roadmap of current technologies and future trends. *Computational Intelligence*, 23(1):69–91.

[60] Franco, C. R.
2017. *Inteligência Artificial*. UNIASSELVI.

[61] Frank, E. and I. H. Witten
1998. Generating accurate rule sets without global optimization. In *Proceedings of the Fifteenth International Conference on Machine Learning*, Pp. 144–151, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[62] Gama, J., A. Ponce, L. Carvalho, K. Faceli, A. C. Lorena, and M. Oliveira
2012. *Extração de Conhecimento de Dados*, 1a. edition. Lisboa: Edições Sílabo Lda.

[63] Gao, C. S. and D. Xu
2006. Conceptual modelling and development of an intelligent agent-assisted decision support system for anti-money laundering. In *Proceedings of the 11th Annual Conference of Asia Pacific Decision Sciences Institute*, Pp. 241–244.

[64] Gao, S. and D. Xu
2010. Real-time exception management decision model (rtemdm): Applications in intelligent agent-assisted decision support in logistics and anti-money laundering domains. In *43rd Hawaii International Conference on System Sciences (HICSS)*, Pp. 1–10.

[65] Gao, S., D. Xu, H. Wang, and Y. Wang
2006. Intelligent anti-money laundering system. In *Service Operations and Logistics, and Informatics, 2006. SOLI '06. IEEE International Conference on*, Pp. 851–856.

[66] Garcia, A. C. B. and J. S. Sichman
2005. *Sistemas Inteligentes - Fundamentos e Aplicações*, chapter 11, Pp. 269–306. Barueri, SP: Manole.

[67] Hamerly, G. and C. Elkan
2002. Alternatives to the k-means algorithm that find better clusterings. In *Proceedings of the Eleventh International Conference on Information and Knowledge Management*, Pp. 600–607, New York, NY, USA. ACM.

[68] Hammoodi, M. S., H. A. A. Essa, and W. A. Hanon
2021. The Waikato Open Source Frameworks (WEKA and MOA) for Machine Learning Techniques. *Journal of Physics: Conference Series*, 1804(1):012133.

[69] Helmy, T. H. E., M. zaki Abd-ElMegied, T. S. Sobh, and K. M. S. Badran
2014. Design of a monitor for detecting money laundering and terrorist financing. *International Journal of Computer Networks and Applications*, 1(1):15–25.

[70] Issicaba, D.
2013. *Block-Oriented Agent-Based Architecture to Support the Power Distribution System Operation - System Design and Enviroment Model*. phdthesis, Institute for Systems and Computer Engineering of Porto, Porto, PT.

[71] Karthikeyan, B.
2020. A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering. *International Journal of Emerging Trends in Engineering Research*, 8(5):1600–1604.

[72] Kashyap, P.
2017. *Machine Learning for Decision Makers*, 1 edition. Apress Berkeley, CA.

[73] Kelleher, J. D., B. M. Namee, and A. D'Arcy
2020. *Fundamentals of Machine Learning for Predictive Data Analytics*. The MIT Press.

[74] Kingdon, J.
2004. Ai fights money laundering. *IEEE Intelligent Systems*, 19(3):87–89.

[75] KPMG
2014. Global anti-money laundering survey. Technical report, KPMG International Corporate. https://www.kpmg.com/KY/en/IssuesAndInsights/ArticlesPublications/PublishingImages/global-anti-money-laundering-survey-v3.pdf, Acessado em 13/01/2016.

[76] Kravari, K. and N. Bassiliades
2015. A survey of agent platforms. *Journal of Artificial Societies and Social Simulation*, 18(1).

[77] Larose, D. T. and C. D. Larose
2014. *Discovering Knowledge in Data: An Introdution to Data Mining*, 2nd edition. John Wiley & Sons. Includes bibliographical references and index.

[78] Laxman, S.
2014. The fight against fraud. *Internal Auditor Online*. https://iaonline.theiia.org/the-fight-against-fraud.

[79] Le-Khac, N.-A. and M.-T. Kechadi
2010. Application of data mining for anti-money laundering detection: A case study. In *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops*, Pp. 577–584, Washington, DC, USA. IEEE Computer Society.

[80] Le-Khac, N.-A., S. Markos, and M. T. Kechadi
2009. Towards a new data mining-based approach for anti-money laundering in an international investment bank. In *Digital Forensics and Cyber Crime - First International ICST Conference (ICDF2C)*, Pp. 77–84, Albany, NY, USA. Springer.

[81] Lei Brasil
1998. Lei nº 9613, de 3 de março de 1998. Diário Oficial da União, http://www.planalto.gov.br/ccivil_03/Leis/L9613.htm.

[82] Lei Portugal
2008. Lei nº 25/2008, de 5 de junho. Diário da República - Série I. Assembleia da República, Disponívelemhttps://www.bportugal.pt/pt-PT/Legislacaoenormas/Documents/Lei25ano2008c.pdf.

[83] Liu, R., X.-l. Qian, S. Mao, and S.-z. Zhu
2011. Research on anti-money laundering based on core decision tree algorithm. In *2011 Chinese Control and Decision Conference (CCDC)*, Pp. 4322–4325.

[84] Liu, X. and P. Zhang
2010. A scan statistics based suspicious transactions detection model for anti-money laundering (aml) in financial institutions. In *Multimedia Communications (Mediacom), 2010 International Conference on*, Pp. 210–213.

[85] Madinger, J.
2012. *Money laundering: a guide for criminal investigators*, 3th edition. Boca Raton, FL: CRC, Taylor & Francis Group.

[86] Mak, M. K., G. T. Ho, and S. Ting
2011. A Financial Data Mining Model for Extracting Customer Behavior. *International Journal of Engineering Business Management*, 3:16.

[87] Mendroni, M. B.
2001. Tópicos essenciais da lavagem de dinheiro. *Revista dos Tribunais*, 90(787):479–489.

[88] Metz, J.
2006. Interpretação de clusters gerados por algoritmos de clustering hierárquico. Dissertação de mestrado em ciências de computação e matemática computacional, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.

[89] Michalski, R., I. Bratko, and A. Bratko
1998. *Machine Learning and Data Mining: Methods and Applications*. Chichester, West Sussex, Eng. New York, NY: John Wiley & Sons.

[90] Monard, M. C. and J. A. Baranauskas
2005. *Conceitos sobre Aprendizado de Máquina*, chapter 4, Pp. 89–114. Barueri, SP: Manole.

[91] MySQL Workbench Community
2014. *MySQL Workbench Refrence Manual*. Oracle.

[92] OIG, Office of Inspector General
2001. Money laundering: Review of the financial crimes enforcement network's use of artificial intelligence to combat money laundering. AUDIT 01-091, Office of Inspector General - The Department of the Treasury, Washington, DC. `https://www.treasury.gov/about/organizational-structure/ig/documents/oig01091.pdf`, Acessado em 04/12/2015.

[93] ONU
1988. United nations convention against illicit traffic in narcotic drugs and psychotropic substances. Convention report, United Nations, Viena. http://www.unodc.org/pdf/convention_1988_en.pdf, Acessado em 25/11/2015.

[94] OTA
1995. *Information Technologies for the Control of Money Laundering*, volume 2 of *Information Technologies for the Control of Money Laundering*, 2 edition. Washington, DC: Congress of the U.S.Office of Technology Assessment, OTA. OTA-ITC-630.

[95] Ovid
2000. *Fasti*. Penguin Books Ltd.

[96] Padgham, L. and M. Winikoff
2004. *Developing Intelligent Agent Systems: A Practical Guide*. Chichester, West Sussex PO19 8SQ, England: John Wiley & Sons Ltd.

[97] Paula, E. L., M. Ladeira, R. N. Carvalho, and T. Marzagão
2016. Deep learning anomaly detection as support fraud investigation in brazilian exports and anti-money laundering. In *15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016, Anaheim, CA, USA, December 18-20, 2016*, Pp. 954–960.

[98] Pham, D. T., S. S. Dimov, and C. D. Nguyen
2005. Selection of k in k-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1):103–119.

[99] Pinto, T., Z. Vale, T. Sousa, I. Praça, G. Santos, and H. Morais
2014. Adaptive learning in agents behaviour: A framework for electricity markets simulation. *Integrated Computer-Aided Engineering*, 21(4):399–415.

[100] Prati, R. C.
2006. *Novas abordagens em aprendizado de máquina para a geração de regras, classes desbalanceadas e ordenação de casos*. PhD thesis, Instituto de Ciências Matemáticas e de Computação, São Carlos.

[101] Quinlan, J. R.
1993. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

[102] Rabkin, N. J.
1998. Money laundering fincen's law enforcement support, regulatory, and international roles. Testimony GAO/T-GGD-98-83, United States General Accounting Office, Washington, D.C. https://books.google.pt/books?id=AmVacntuOVsC.

[103] Rajput, Q., N. S. Khan, A. S. Larik, and S. Haider
2014. Ontology based expert-system for suspicious transactions detection. *Computer and Information Science*, 7(1):103–114.

[104] Rezende, S. O., J. B. Pugliesi, E. A. Melanda, and M. F. de Paula
2005. *Mineração de Dados*, chapter 12, Pp. 307–336. Barueri, SP: Manole Ltda.

[105] Rodriguez-Arias, A., B. Guijarro-Berdinas, and N. Sanchez-Marono
2020. A FIPA-ACL based communication utility for unity. In *2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*. IEEE.

[106] Rousseeuw, P.
1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, 20(1):53–65.

[107] Russell, S. and P. Norvig
2020. *Artificial Intelligence: A Modern Approach*, 4rd edition. Upper Saddle River, NJ, USA: Prentice Hall Press.

[108] Sabau, A. S.
2012. Survey of clustering based financial fraud detection research. *Informatica Economica*, 16(1).

[109] Salles, C.
2014. *Dinheiro História, Mitos & Crenças - o sentido e significado dos valores*. e-Book.

[110] Satula, B.
2010. *Branqueamento de Capitais*. Lisboa: Universidade Católica Editora. 144p.

[111] Scholkopf, B., J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson
2001. Estimating the support of a high-dimensional distribution. *Neural Computing*, 13(7):1443–1471.

[112] Schott, P. A.
2006. *Reference Guide to Anti-Money Laundering and Combating the Financing of Terrorism: Second Edition and Supplement on Special Recommendation IX*, second edition. Washington DC: The World Bank and The International Monetary Fund.

[113] Senator, T. E., H. G. Goldberg, J. Wooton, M. A. Cottini, A. F. U. Khan, C. D. Klinger, W. M. Llamas, M. P. Marrone, and R. W. H. Wong
1995. The financial crimes enforcement network ai system (fais) identifying potential money laundering from reports of large cash transactions. *AI Magazine*, 16(4):21–39.

[114] Shehory, O. and A. Sturm
2014. *Agent-Oriented Software Engineering: Reflections on Architectures, Methodologies, Languages, and Frameworks*, SpringerLink : Bücher. Springer Berlin Heidelberg.

[115] Sichman, J. S. and H. Coelho
2014. *Negotiation and Argumentation in Multi-Agent Systems: Fundamentals, Theories, Systems and Applications*, chapter 1, Pp. 3–29. Sharjah, U.A.E: Bentham Science Publishers Ltd.

[116] Signoretti, A.
2012. *Agentes Inteligentes com Foco de Atenção Afetivo em Simulações Baseadas em Agentes*. phdthesis, Centro de Tecnologia, UFRN, Natal, RN.

[117] Tai, C.-H. and T.-J. Kan
2019. Identifying money laundering accounts. In *2019 International Conference on System Science and Engineering (ICSSE)*, Pp. 379–382. IEEE.

[118] Tan, P.-N., M. Steinbach, V. Kumar, and A. Karpatne
2019. *Introduction to Data Mining, Global Edition*, 2a edition. United Kingdom: Pearson Education Limited.

[119] Tang, J. and J. Yin
2005. Developing an intelligent data discriminating system of anti-money laundering based on svm. In *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, volume 6, Pp. 3453–3457.

[120] The World Bank
2017. World databank. Global Findex (Global Financial Inclusion Database), `http://databank.worldbank.org/data/databases.aspx`. Acessado em 09/01/2021.

[121] UNODC
2019. United Nations Office on Drugs and Crime - Annual Report 2018. Online. `https://www.unodc.org/unodc/en/about-unodc/annual-report.html`, Acessado em jan/2020.

[122] USDT
2018. National Money Laundering Risk Assessment 2018. techreport 2018 NMLRA, United State of Department of the Treasury, NW Washington, D.C.

[123] Wagner, S. and D. Wagner
2007. Comparing clusterings – an overview. Technical Report 2006-04, "Universität Karlsruhe (TH)". `http://digbib.ubka.uni-karlsruhe.de/volltexte/1000011477`.

[124] Weiss, S. and N. Indurkhya
1998. *Predictive Data Mining: A Practical Guide*, The Morgan Kaufmann Series in Data Management Systems. Elsevier Science.

[125] Wilson, D. R. and T. R. Martinez
1997. Improved heterogeneous distance functions. *J. Artif. Int. Res.*, 6(1):1–34.

[126] Winikoff, M.
2007. Defining syntax and providing tool support for agent uml using a textual notation. *Int. J. Agent-Oriented Software Engineering*, 1(2):123–144. `http://dx.doi.org/10.1504/IJAOSE.2007.014406`, Acessado em 20/01/2016.

[127] Winikoff, M. and L. Padgham
2013. *Agent-Oriented Software Engineering*, chapter 15, Pp. 695–758. London, England: The MIT Press, 2nd edition.

[128] Wooldridge, M.
2009. *An Introduction to Multiagent Systems*, 2nd edition. Chichester, UK: Wiley Publishing.

[129] Wooldridge, M.
2013. *Multiagent systems*, chapter 1, Pp. 3–50. Cambridge, Massachusetts London, England: The MIT Press.

[130] Wooldridge, M. and N. R. Jennings
1995. Intelligent agents: theory and practice. *The Knowledge Engineering Review*, 10(2):115–152.

[131] Xuan, L. and Z. Pengzhu
2007. An agent based anti-money laundering system architecture for financial supervision. In *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*, Pp. 5472–5475.

[132] Zhang, Z. M., J. J. Salerno, and P. S. Yu
2003. Applying data mining in investigating money laundering crimes. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, Pp. 747–752, New York, NY, USA. ACM.